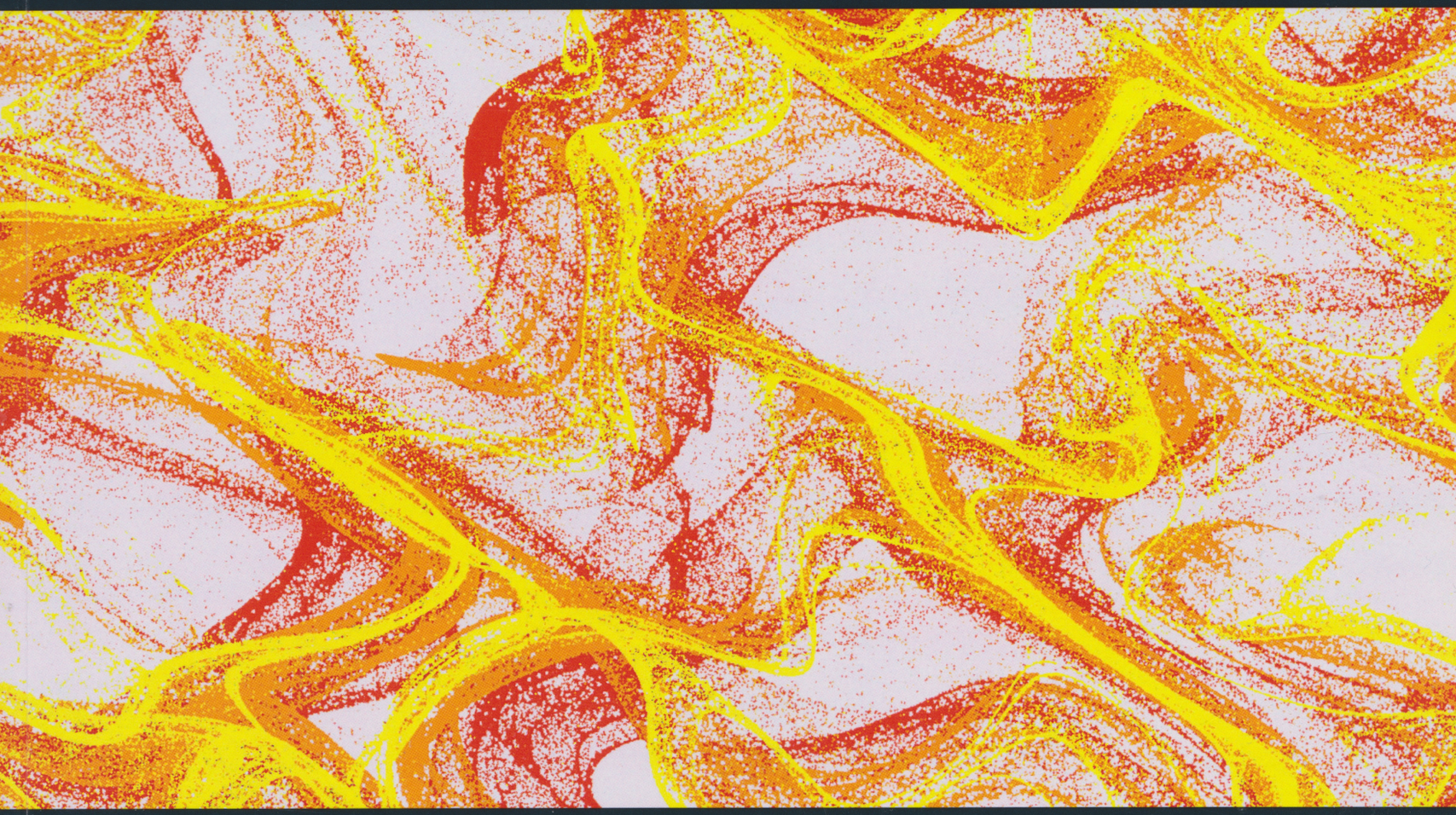
The Open University

MST224

Mathematical methods

Handbook



The Open University

MST224

Mathematical methods

Handbook

This publication forms part of an Open University module. Details of this and other Open University modules can be obtained from Student Recruitment, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)300 303 5303; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2013. Second edition 2016.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Halstan & Co. Ltd, Amersham, Bucks.

SUP 05024 3

2.1

Contents

Introduction

This handbook is a reference that you can take into the MST224 exam, and it may be valuable when you start to apply your knowledge in other modules. It will be more effective if you are already familiar with it before you sit the exam, and we suggest that you consult it when you attempt the assignments in the module. This handbook is not designed as a teaching document, and reading it is not a substitute for studying the module units.

The first few sections consist of general mathematical reference material, largely based on the topics in Unit 1. It is not intended to be a self-contained or logically complete account of basic mathematics; it is just a set of definitions and results. Some of these results are used repeatedly in the module. You will also find some additional results or definitions that are not covered in Unit 1, but are useful to have in a mathematical reference booklet.

The later sections of the handbook are brief summaries of the units, emphasising the most important results.

1 Notation

1.1 Greek alphabet

α	A	alpha	ι	I	iota	ρ	P	rho
β	B	beta	κ	K	kappa	σ	Σ	sigma
γ	Γ	gamma	λ	Λ	lambda	τ	T	tau
δ	Δ	delta	μ	M	mu	υ	Υ	upsilon
ε	E	epsilon	ν	N	nu	ϕ	Φ	phi
ζ	Z	zeta	ξ	Ξ	xi	χ	X	chi
η	H	eta	o	O	omicron	ψ	Ψ	psi
θ	Θ	theta	π	Π	pi	ω	Ω	omega

1.2 Symbols

$=$	is equal to
$\neq$	is not equal to
$\simeq$	is approximately equal to
$\pm$	plus or minus
$\mp$	minus or plus
$<$	less than
$\leq$	less than or equal to
$>$	greater than
$\geq$	greater than or equal to
$\sqrt{\ }$	positive square root
e	the number 2.718 28 ...
π	the number 3.141 59 ...
∞	infinity
$\mathbb{Z}$	the integers
$\mathbb{R}$	the real numbers
$\mathbb{C}$	the complex numbers

1.3 Limits and sums

An ordered list of numbers $x_0, x_1, x_2, \ldots, x_n, \ldots$ is said to **converge** to the **limit** x if successive terms in the ordered list are better and better approximations to x. We write '$x_n \to x$ as $n \to \infty$' or '$\lim_{n\to\infty} x_n = x$'.

Given numbers $a_1, a_2, \ldots, a_n$, we define

$$\sum_{i=1}^{n} a_i \quad \text{to mean} \quad a_1 + a_2 + \cdots + a_{n-1} + a_n.$$

2 Numbers

2.1 Real numbers

The **integers** are the positive and negative whole numbers, together with zero. Non-integer numbers that can be expressed exactly as fractions are called **rational numbers**; those that cannot be so expressed, such as $\sqrt{2}$, e and π, are called **irrational**. The collection of the rational numbers (including the integers) and the irrational numbers is called the set of **real numbers**.

When a real number is expressed in decimal notation, if it is approximated then the approximation can be given to so many **decimal places**, or to so many **significant figures**. For example, 1.4142 is the approximation to $\sqrt{2}$ to four decimal places and five significant figures, while 0.000 000 342 is given to nine decimal places but three significant figures, and 342 000 has no decimal places but at least three significant figures. The process of

reducing the number of decimal places or significant figures to which a number is expressed is referred to as **rounding**. To round a given number to n decimal places or n significant figures, take the number expressed to n decimal places or n significant figures that is closest to the given number, where it is conventional to round the digit 5 up. (Be aware that other conventions for rounding exist, and that computer programs and calculators do not always use this convention.)

A number given in the form $\pm b \times 10^c$, where $1 \leq b < 10$ and c is an integer, is said to be in **scientific notation**. For example, the number 342 000 can be expressed as 3.42×10^5 in scientific notation.

For any two real numbers a and b with $a < b$, we write:

$[a, b]$ for the set of all real numbers x such that $a \leq x \leq b$;

$[a, b)$ for the set of all real numbers x such that $a \leq x < b$;

$(a, b]$ for the set of all real numbers x such that $a < x \leq b$;

(a, b) for the set of all real numbers x such that $a < x < b$.

These sets of numbers are called **intervals**. The interval $[a, b]$ is a **closed** interval, (a, b) is an **open** interval, and the two others are **half-open** intervals. For practical problems the distinction between open and closed intervals is rarely significant. If we know that a real number x is 1.274 to three decimal places, then x lies in the interval $[1.2735, 1.2745)$.

2.2 Complex numbers

A **complex number** z is written in **Cartesian form** as $z = a + bi$, where a and b are real numbers, and $i^2 = -1$. We refer to a as the **real part** of z, written $\mathrm{Re}(z)$, and to b as the **imaginary part** of z, written $\mathrm{Im}(z)$.

Complex numbers can be added, e.g.

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

or multiplied, e.g.

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

These formulas make use of the ordinary rules of algebra, together with the relation $i^2 = -1$.

The **complex conjugate** of $z = a + bi$ is $\overline{z} = a - bi$. Note that

$$z\overline{z} = (a + bi)(a - bi) = a^2 + b^2$$

is a positive real number (unless $a = b = 0$). The **modulus** of z is the number $|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$.

The notation z^* is also used in some texts for the complex conjugate of z.

$\overline{z}$ and z^* are read as 'z bar' and 'z star'.

To calculate a quotient of complex numbers, multiply top and bottom by the complex conjugate of the bottom, e.g.

$$\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \left(\frac{ac + bd}{c^2 + d^2}\right) + \left(\frac{bc - ad}{c^2 + d^2}\right) i,$$

which expresses the quotient in the form $p + qi$ with p and q real.

Polar coordinates are discussed in Subsection 5.2.

The **Argand diagram** is a representation of complex numbers as points in a plane, where the complex number $a + bi$ is represented by the point with Cartesian coordinates (a, b). A point can also be represented in polar coordinates as (r, θ) and related to its Cartesian coordinates by

$$a = r\cos\theta, \quad b = r\sin\theta.$$

The angle coordinate θ is referred to as an **argument** of z, written $\arg(z)$, and the unique value of θ in the range $-\pi < \theta \leq \pi$ is referred to as the **principal value of the argument**, written $\mathrm{Arg}(z)$.

The multiplication of complex numbers in polar form is given by the rule $(r, \theta) \times (s, \phi) = (rs, \theta + \phi)$. So powers of a complex number can be expressed as $(r, \theta)^n = (r^n, n\theta)$, for n a positive integer. The special case when $r = 1$ is known as **de Moivre's theorem**:

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta).$$

Euler's formula is

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

This extends to any complex number $a + bi$ as

$$e^{a+bi} = e^a e^{bi} = e^a(\cos b + i\sin b).$$

The **exponential form** of a complex number $z = r(\cos\theta + i\sin\theta)$ is

$$z = re^{i\theta}.$$

This form is useful for multiplying, dividing and taking powers of complex numbers.

3 Functions and graphs

3.1 Functions

A **variable** is a quantity, represented by a symbol, that can vary over a set of values. If its value does not vary, then it is a **constant**.

Any expression or formula that involves a variable x, and whose value is uniquely determined by the value of x, is called a **function** of x.

If a variable y is a function of x (i.e. if y is equal to a function of x), then we call x the **independent variable** and y the **dependent variable**, and we may write $y = y(x)$. Here $y(x)$ stands for the function of x (i.e. for the formula involving x).

If f and g are two functions, then their **sum** is a function $f + g$ defined by

$$(f + g)(x) = f(x) + g(x) \quad \text{(for all } x\text{)}.$$

Moreover, if A and B are any two numbers, then the function $Af + Bg$ is defined by

$$(Af + Bg)(x) = A\,f(x) + B\,g(x) \quad \text{(for all } x\text{)}.$$

The function $f(g(x))$ is called the **composite function** or **composition** of the functions f and g.

The **graph** of a function $f(x)$ is the curve in the xy-plane whose equation is $y = f(x)$.

A **continuous** function is one whose graph has no breaks or jumps in it, i.e. it can be drawn without lifting your pen from the paper.

A **constant function** $f(x)$ is one that assigns the same value to any input. Its graph is a straight line parallel to the x-axis. A special case is the **zero function**, which assigns the value 0 to any input.

A **linear function** is one having the form $a_1x + a_0$ (with $a_1 \neq 0$), where a_1 and a_0 are constants. Its graph is a straight line with **slope** a_1 and ***y*-intercept** a_0.

A **quadratic function** is one having the form $a_2x^2 + a_1x + a_0$ (with $a_2 \neq 0$), where a_2, a_1, a_0 are constants. Its graph is a parabola, similar in shape to the one shown in Subsection 5.3 if $a_2 > 0$, but the other way up if $a_2 < 0$.

A **cubic function** is one having the form $a_3x^3 + a_2x^2 + a_1x + a_0$ (with $a_3 \neq 0$), where a_3, a_2, a_1, a_0 are constants.

3.2 Polynomials

Linear, quadratic and cubic functions are all particular examples of **polynomial** functions, or simply polynomials. An ***n*th-order** polynomial, or polynomial of **degree *n***, is a function of the form

$$a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0,$$

where n is a positive integer, x is a variable and $a_0, a_1, \ldots, a_n$ are constants with $a_n \neq 0$. A linear polynomial has $n = 1$ (or $n = 0$), a quadratic polynomial has $n = 2$, and a cubic polynomial has $n = 3$.

The **roots** of a polynomial $p(x)$ are the solutions of the equation $p(x) = 0$. Every polynomial of degree n can be written as a product of a_n and n factors of the form $x - c_k$ $(k = 1, 2, \ldots, n)$, with each c_k a complex number (which may be real). Each of these factors corresponds to a root $x = c_k$ of the polynomial. If a factor $x - c$ occurs more than once, then the root $x = c$ is a **repeated root**; repeated roots are also sometimes referred to as **equal roots** or **coincident roots**.

The roots of a quadratic equation $ax^2 + bx + c = 0$, $a \neq 0$, are given by the **formula method** as

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The quantity $b^2 - 4ac$ is referred to as the **discriminant** of the quadratic equation.

To **factorise** a polynomial is to express it as a product of two or more polynomials of lower degree. For example, the **difference of two squares** $x^2 - a^2$ factorises as

$$x^2 - a^2 = (x - a)(x + a),$$

and the **perfect square** $x^2 + 2ax + a^2$ factorises as

$$x^2 + 2ax + a^2 = (x + a)(x + a) = (x + a)^2.$$

3.3 Exponentials and logarithms

A function of the form $y = ba^x$, where a and b are non-zero constants (with $a > 0$ and $a \neq 1$), and x is real, is said to exhibit **exponential behaviour**. In a^x, a is referred to as the **base** and x as the **exponent** (or **index** or **power**). Properties of such a function include

$$a^0 = 1,$$
$$a^{-x} = 1/a^x,$$
$$a^x \times a^y = a^{x+y},$$
$$a^x / a^y = a^{x-y},$$
$$(a^x)^y = a^{xy} = (a^y)^x.$$

Some texts refer to $y = ba^x$ as *an* exponential function; this is not to be confused with *the* exponential function, $\exp x$.

The function e^x, where $e = 2.718\,28\ldots$, is referred to as the **exponential function**. It is also written as $\exp x$.

The **natural logarithm function** $\ln x$ is defined to be the **inverse function** of the exponential function $\exp x$, i.e. each reverses the effect of the other, so that

$$\ln(\exp x) = x \quad \text{for all real } x,$$
$$\exp(\ln x) = x \quad \text{for all real } x > 0.$$

In other words, if $e^y = x$, then $y = \ln x$, and vice versa.

The natural logarithm function $\ln x$, for $x > 0$, has the properties

$$\ln 1 = 0,$$
$$\ln(1/x) = -\ln x,$$
$$\ln(xy) = \ln x + \ln y, \quad y > 0,$$
$$\ln(x/y) = \ln x - \ln y, \quad y > 0,$$
$$\ln x^y = y \ln x.$$

Any function $y = ba^x$ can be written in the form $y = be^{kx}$, where $k = \ln a$.

Another logarithm function is $\log_{10} x$, for $x > 0$, where $y = \log_{10} x$ if $10^y = x$ (and vice versa). The properties given above for ln also hold for $\log_{10}$.

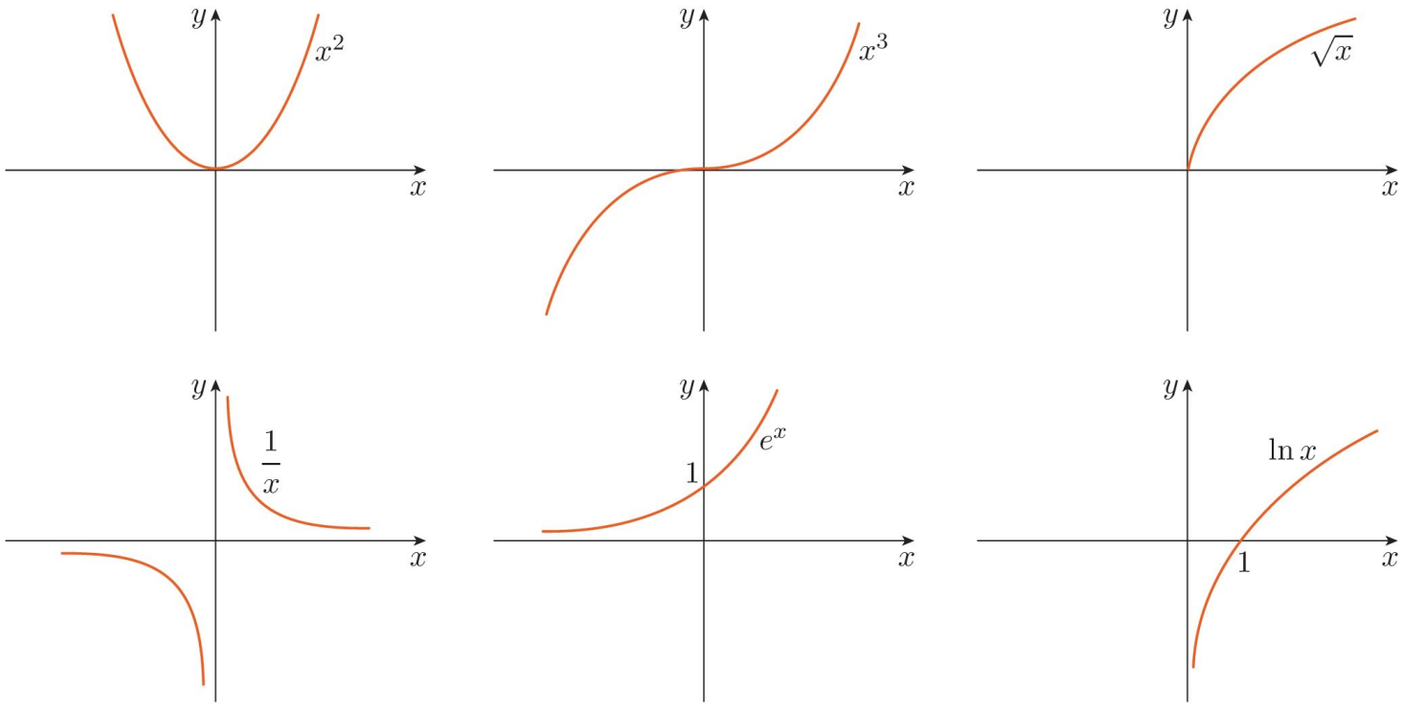
Although they are not used in this module, we note for general reference purposes that the **hyperbolic functions** $\sinh x$, $\cosh x$ and $\tanh x$ are defined as combinations of exponential functions:

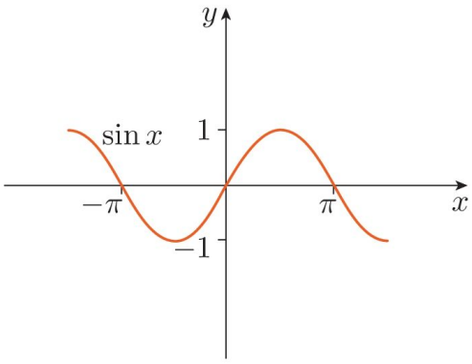
$$\sinh x = \tfrac{1}{2}(e^x - e^{-x}),$$
$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}),$$
$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

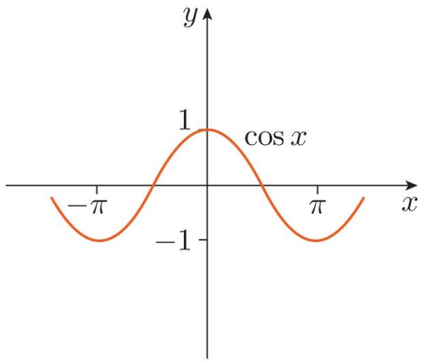
The inverses of these functions are

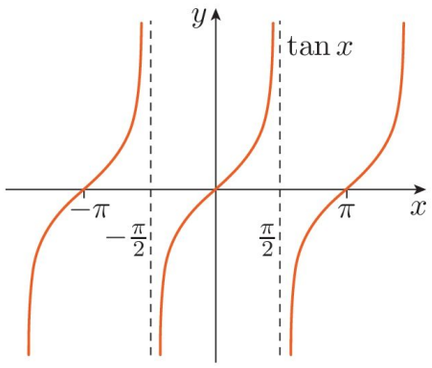
$$\operatorname{arcsinh} x = \ln(x + \sqrt{x^2 + 1}),$$
$$\operatorname{arccosh} x = \ln(x + \sqrt{x^2 - 1}), \quad x \geq 1,$$
$$\operatorname{arctanh} x = \tfrac{1}{2} \ln\left(\frac{1+x}{1-x}\right), \quad |x| < 1.$$

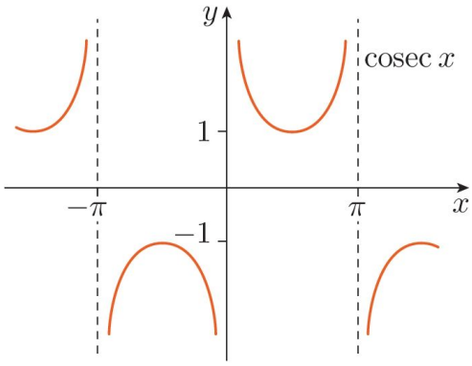
3.4 Graphs of some common functions

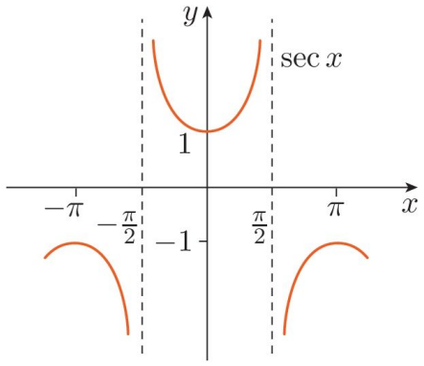


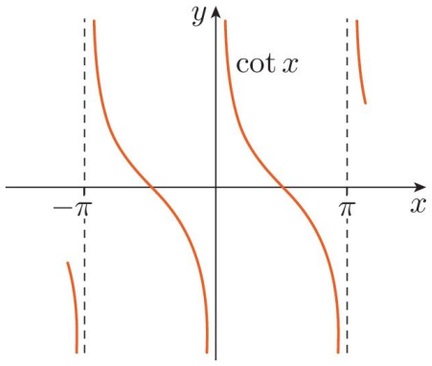

y
sin x
1
−π
π
x
−1

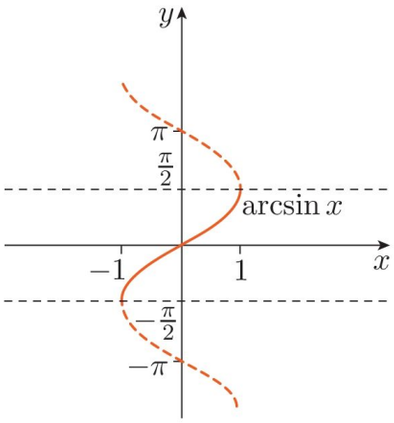

y
1
cos x
−π
π
x
−1

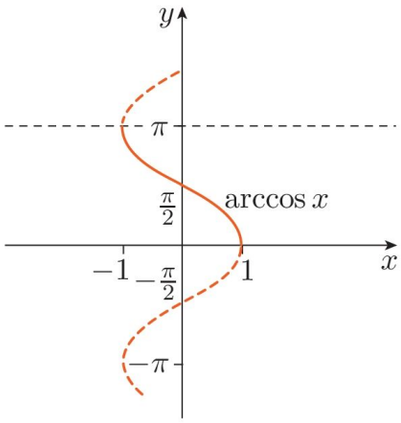

y
tan x
−π
−π/2
π/2
π
x

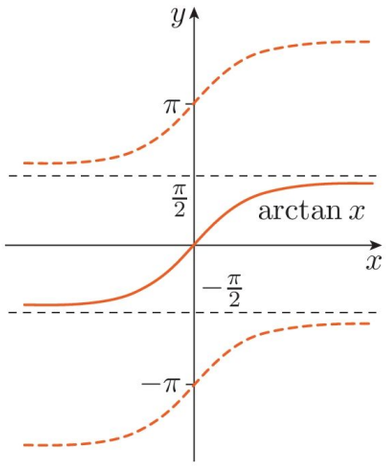

y
cosec x
1
−π
π
x
−1

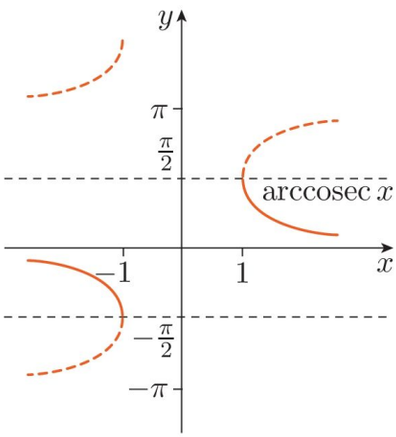

y
sec x
1
−π
−π/2
π/2
π
x
−1

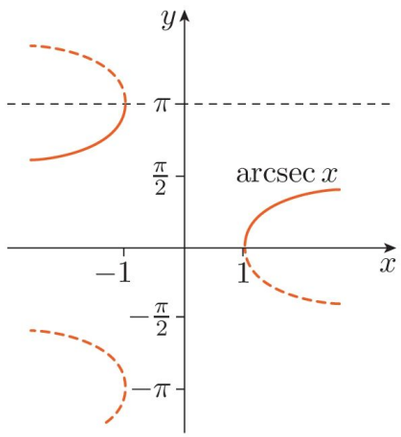

y
cot x
−π
π
x

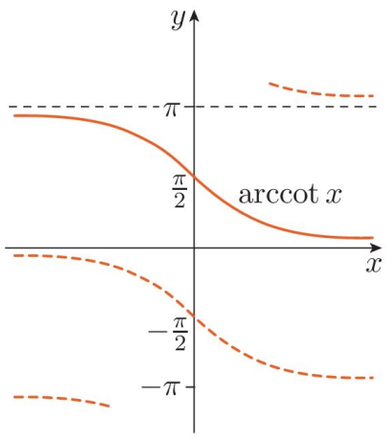

y
π
π/2
arcsin x
−1
1
x
−π/2
−π


y
π
π/2
arccos x
−1
−π/2
1
x
−π


y
π
π/2
arctan x
x
−π/2
−π


y
π
π/2
arccosec x
−1
1
x
−π/2
−π


y
π
π/2
arcsec x
−1
1
x
−π/2
−π


y
π
π/2
arccot x
x
−π/2
−π

4 Trigonometry

4.1 Radians and degrees

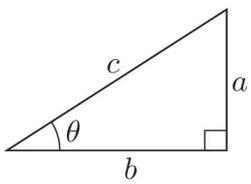
In this module we usually measure angles in **radians** rather than degrees. There are 2π radians in a full circle, corresponding to $360°$, so 1 radian is $(180/\pi)° \simeq 57°$. An advantage of working in radians is the simplicity of the formula for the arc length subtended by an angle in a circle of radius r: the length of the arc subtended by an angle of θ radians is simply $r\theta$. The following radian measures of standard angles are worth knowing:

- a right angle is $\frac{\pi}{2}$ radians
- the angles of an equilateral triangle are $\frac{\pi}{3}$ radians.

An angle is **acute** if its radian measure lies between 0 and $\frac{\pi}{2}$, and **obtuse** if its radian measure lies between $\frac{\pi}{2}$ and π.

4.2 Trigonometric functions and their inverses

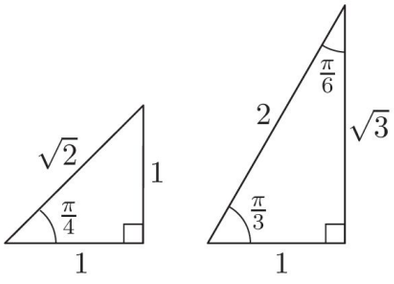
For an acute angle θ, the values of trigonometric functions are related to the ratios of lengths of the sides of a triangle as follows.



Function	Definition for acute angles in terms of triangle shown	Definition in terms of sin and cos
$\sin\theta$	$\frac{a}{c}$	
$\cos\theta$	$\frac{b}{c}$	
$\tan\theta$	$\frac{a}{b}$	$\frac{\sin\theta}{\cos\theta}$
$\cot\theta$	$\frac{b}{a}$	$\frac{\cos\theta}{\sin\theta}$
$\sec\theta$	$\frac{c}{b}$	$\frac{1}{\cos\theta}$
$\operatorname{cosec}\theta$	$\frac{c}{a}$	$\frac{1}{\sin\theta}$

Inverse function	Definition
$\arcsin x$	$= \theta$ where $\sin\theta = x$
$\arccos x$	$= \theta$ where $\cos\theta = x$
$\arctan x$	$= \theta$ where $\tan\theta = x$
$\operatorname{arccot} x$	$= \theta$ where $\cot\theta = x$
$\operatorname{arcsec} x$	$= \theta$ where $\sec\theta = x$
$\operatorname{arccosec} x$	$= \theta$ where $\operatorname{cosec}\theta = x$

4.3 Two useful triangles



From the two triangles in the margin it can be seen that:

$$\sin\tfrac{\pi}{6} = \tfrac{1}{2}, \quad \cos\tfrac{\pi}{6} = \tfrac{\sqrt{3}}{2}, \quad \tan\tfrac{\pi}{6} = \tfrac{1}{\sqrt{3}},$$
$$\operatorname{cosec}\tfrac{\pi}{6} = 2, \quad \sec\tfrac{\pi}{6} = \tfrac{2}{\sqrt{3}}, \quad \cot\tfrac{\pi}{6} = \sqrt{3};$$
$$\sin\tfrac{\pi}{4} = \tfrac{1}{\sqrt{2}}, \quad \cos\tfrac{\pi}{4} = \tfrac{1}{\sqrt{2}}, \quad \tan\tfrac{\pi}{4} = 1,$$
$$\operatorname{cosec}\tfrac{\pi}{4} = \sqrt{2}, \quad \sec\tfrac{\pi}{4} = \sqrt{2}, \quad \cot\tfrac{\pi}{4} = 1;$$
$$\sin\tfrac{\pi}{3} = \tfrac{\sqrt{3}}{2}, \quad \cos\tfrac{\pi}{3} = \tfrac{1}{2}, \quad \tan\tfrac{\pi}{3} = \sqrt{3},$$
$$\operatorname{cosec}\tfrac{\pi}{3} = \tfrac{2}{\sqrt{3}}, \quad \sec\tfrac{\pi}{3} = 2, \quad \cot\tfrac{\pi}{3} = \tfrac{1}{\sqrt{3}}.$$

Other values of the trigonometric functions worth remembering are:

$$\sin 0 = 0, \quad \cos 0 = 1, \quad \tan 0 = 0;$$
$$\sin\tfrac{\pi}{2} = 1, \quad \cos\tfrac{\pi}{2} = 0;$$
$$\sin\pi = 0, \quad \cos\pi = -1, \quad \tan\pi = 0;$$
$$\sin\tfrac{3\pi}{2} = -1, \quad \cos\tfrac{3\pi}{2} = 0.$$

4.4 Trigonometric identities

Pythagoras's theorem states that for any right-angled triangle, if c is the length of the hypotenuse (the side opposite the right angle) and a and b are the lengths of the other two sides, then

$$c^2 = a^2 + b^2.$$

This leads to the following trigonometric identities:

$$\sin^2\theta + \cos^2\theta = 1,$$
$$\tan^2\theta + 1 = \sec^2\theta,$$
$$1 + \cot^2\theta = \operatorname{cosec}^2\theta.$$

Addition formulas

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta,$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta,$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta,$$
$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta,$$
$$\tan(\alpha+\beta) = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta},$$
$$\tan(\alpha-\beta) = \frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta};$$

$$\sin\alpha\cos\beta = \tfrac{1}{2}\sin(\alpha+\beta) + \tfrac{1}{2}\sin(\alpha-\beta),$$
$$\cos\alpha\sin\beta = \tfrac{1}{2}\sin(\alpha+\beta) - \tfrac{1}{2}\sin(\alpha-\beta),$$
$$\cos\alpha\cos\beta = \tfrac{1}{2}\cos(\alpha+\beta) + \tfrac{1}{2}\cos(\alpha-\beta),$$
$$\sin\alpha\sin\beta = \tfrac{1}{2}\cos(\alpha-\beta) - \tfrac{1}{2}\cos(\alpha+\beta).$$

In particular, these formulas give

$$\sin(\alpha+2\pi) = \sin\alpha, \quad \cos(\alpha+2\pi) = \cos\alpha, \quad \tan(\alpha+\pi) = \tan\alpha;$$
$$\sin(-\alpha) = -\sin\alpha, \quad \cos(-\alpha) = \cos\alpha, \quad \tan(-\alpha) = -\tan\alpha.$$

Double-angle formulas

$$\sin 2\alpha = 2\sin\alpha\cos\alpha,$$
$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 1 - 2\sin^2\alpha = 2\cos^2\alpha - 1,$$
$$\tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha},$$
$$\sin^2\alpha = \tfrac{1}{2}(1-\cos 2\alpha),$$
$$\cos^2\alpha = \tfrac{1}{2}(1+\cos 2\alpha).$$

Cosines of related angles

$$\cos(\tfrac{\pi}{2}-\alpha) = \sin\alpha, \quad \cos(\tfrac{\pi}{2}+\alpha) = -\sin\alpha,$$
$$\cos(\pi-\alpha) = -\cos\alpha, \quad \cos(\pi+\alpha) = -\cos\alpha.$$

4.5 General sinusoidal functions

A **sinusoidal function** or **sinusoid** is a function $x(t)$ of time t of the form

$$x = x_0 + A\cos(\omega t+\phi) = x_0 + A\sin(\omega t+\psi),$$

where x_0 is a constant, A is a positive constant called the **amplitude**, ω is a positive constant called the **angular frequency**, and ϕ and ψ are constants called **phase constants**.

A sinusoidal function oscillates between $x_0 - A$ and $x_0 + A$, repeating the same pattern of oscillations through each time interval of length $2\pi/\omega$, known as the **period** of the function. For these reasons, sinusoidal functions are examples of **oscillatory** functions and of **periodic** functions.

The phase constants in the two forms of the sinusoidal function are related according to $\phi = \frac{3\pi}{2} + \psi$ (or equivalently $\phi = \psi - \frac{\pi}{2}$).

Alternative forms of the sinusoidal functions are given by

$$x = x_0 + A\cos(\omega t + \phi) = x_0 + B\cos(\omega t) + C\sin(\omega t), \tag{1}$$

$$x = x_0 + A\sin(\omega t + \psi) = x_0 + D\sin(\omega t) + E\cos(\omega t). \tag{2}$$

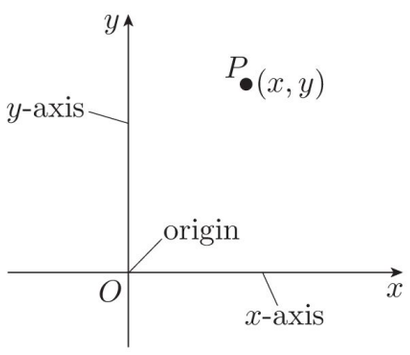
In equation (1) we have $B = A\cos\phi$ and $C = -A\sin\phi$. Solving these equations gives $A = \sqrt{B^2 + C^2}$ and $\phi = \arccos(B/A)$.

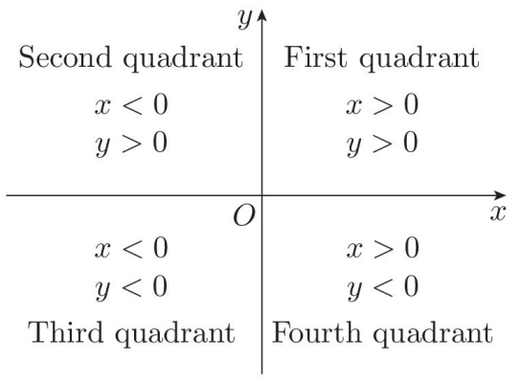
In equation (2) we have $D = A\cos\psi$ and $E = A\sin\psi$. Solving these equations gives $A = \sqrt{D^2 + E^2}$ and $\psi = \arccos(D/A)$.

5 Geometry

5.1 Cartesian coordinates

The **Cartesian coordinates** (x, y) of a point P in a plane specify the position of that point relative to two perpendicular axes, the ***x*-axis** (or **horizontal axis**) and ***y*-axis** (or **vertical axis**), which meet at a point O called the **origin**, with Cartesian coordinates $(0, 0)$. The directions of the axes indicate increasing numerical values for the x- and y-coordinates. Values of x to the right of the y-axis are positive, and those to the left are negative; similarly, values of y above the x-axis are positive, and those below are negative. The four parts into which a plane is divided by Cartesian coordinate axes are known as **quadrants** of the plane. A plane on which Cartesian coordinate axes have been specified is often referred to as the ***xy*-plane**.



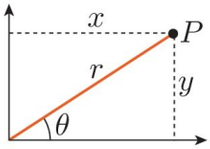


5.2 Polar coordinates

The point P whose polar coordinates are (r, θ) has Cartesian coordinates (x, y) where

$$x = r\cos\theta, \quad y = r\sin\theta.$$

The value of r is always positive (except at the origin, where it is zero). For a given point P, the value of θ is not unique: we can add or subtract any integer multiple of 2π and obtain another value for θ that describes the same point. The value of θ satisfying $-\pi < \theta \leq \pi$ is called the **principal value** of θ.



The relationship between polar and Cartesian coordinates

5.3 Plane figures and curves

A closed plane figure with straight sides is called a **polygon**. A polygon with 3 sides is a **triangle**, one with 4 sides is a **quadrilateral**, one with 5 sides is a **pentagon**, one with 6 sides is a **hexagon**, and in general one with n sides is called an ***n*-gon**.

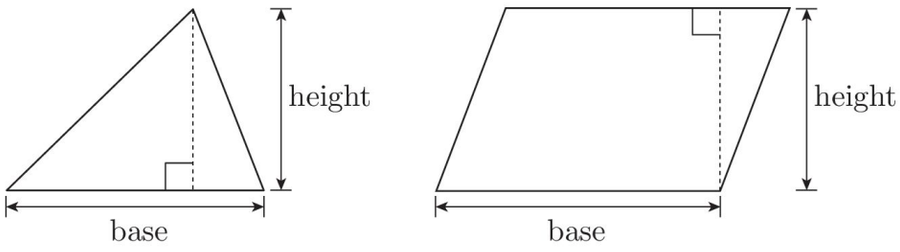
A polygon is said to be **regular** if all its sides have equal length and all its angles are equal. A regular triangle is referred to as an **equilateral triangle**, and a regular quadrilateral is a **square**.

An **isosceles triangle** is one with two sides of equal length (or equivalently with two equal angles). A **right-angled triangle** is one in which one angle is a right angle. The **angle sum of a triangle** is π radians (180°).

A **parallelogram** is a quadrilateral with opposite sides parallel.
A **rectangle** is a parallelogram all of whose angles are right angles.
A **square** is a rectangle all of whose sides have equal length. The **angle sum of a quadrilateral** is 2π radians (360°).

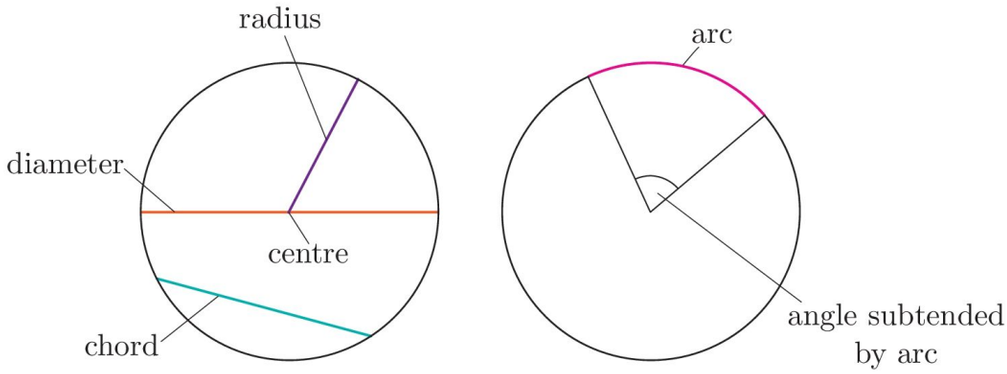
The **area of a triangle** $= \frac{1}{2} \times$ base $\times$ height.

The **area of a parallelogram** $=$ base $\times$ height.



The areas of a triangle and parallelogram

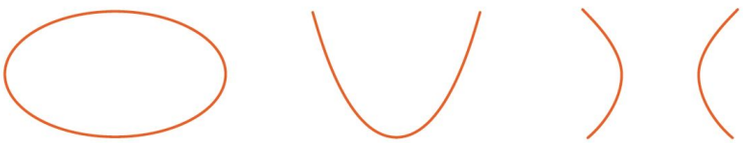
A **circle** is a set of points in a plane that are a constant distance from a fixed point in the plane. The fixed point is the **centre** of the circle, and the constant distance is its **radius**. If a straight line cuts a circle at two points, then the segment of that straight line within the circle is known as a **chord** of the circle. The length of a chord that passes through the centre of a circle is the **diameter** of the circle. The terms *diameter* and *radius* are also used to refer to a chord through the centre of a circle and to a straight line from a point on the circle to its centre, respectively.



Each continuous segment of a circle is known as an **arc** of the circle; the angle made at the centre of a circle by two radii drawn from the ends of the arc is known as the **angle subtended by the arc**. (In such circumstances the arc itself is sometimes called the arc subtended by the angle.)

The distance around a circle is known as its **circumference**, and for a circle of radius r this is given by $2\pi r$. The **area of a circle** of radius r is πr^2. The **arc length** of an arc of a circle of radius r subtended by an angle θ radians is $r\theta$. The **area of a sector** of a circle of radius r subtended by an angle θ radians is $\frac{1}{2}r^2\theta$.

The **equation of a circle** in the xy-plane with centre (a, b) and radius r is $(x-a)^2 + (y-b)^2 = r^2$. Other curves in the xy-plane that can be represented by quadratic formulas are the **ellipse**, the **parabola** and the **hyperbola**, examples of which are shown below.



Ellipse Parabola Hyperbola

A straight line between two distinct points on a curve is known as a **chord** of the curve. A straight line that just touches a curve is known as a **tangent** to that curve at the point where it touches.

6 Differentiation

6.1 Notation and terminology

If $f(x)$ is a function, then its **derived function** or **derivative** $f'(x)$ is defined by

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

The process of calculating $f'(x)$ from $f(x)$ is called **differentiation** of $f(x)$ with respect to x. Differentiation with respect to x can also be denoted by the symbol $\dfrac{d}{dx}$ written to the left of the expression or variable being differentiated, so that, for example,

$$\frac{d}{dx}(f(x)) \quad \text{and} \quad \frac{df(x)}{dx} \quad \text{both mean} \quad f'(x).$$

If $y = f(x)$, then $f'(x)$ can also be written as y' or $\dfrac{dy}{dx}$, where to save space we often print dy/dx in place of $\dfrac{dy}{dx}$.

When the independent variable is t (time), we often use a dot to indicate a derivative, so that $\dot{u}$ means the same thing as $u'(t)$ or du/dt. If x is the position of a particle along an axis, then dx/dt or $\dot{x}$ represents the **velocity** along the axis, and d^2x/dt^2 or $\ddot{x}$ represents the **acceleration** along the axis.

The notation $f'(x)$ is referred to as **function** notation, dy/dx as **Leibniz** notation, and $\dot{u}$ as **Newton's** or **Newtonian** notation.

The derivative of a derivative is called a **second derivative**. For example, the second derivative of the function $f(x)$, denoted by $f''(x)$, is the derivative of $f'(x)$. If $y = f(x)$, then the second derivative is also written as y'' or d^2y/dx^2. If u is a function of t, then its second derivative can be written as $\ddot{u}$.

Third and **higher derivatives** are defined analogously. The nth derivative of f is denoted by $f^{(n)}$ or, if $y = f(x)$, by d^ny/dx^n; n is referred to as the **order** of the derivative. The prime and dot notations are not used for higher derivatives, except that f''' is sometimes used in place of $f^{(3)}$.

A **complex-valued** function $f(x) = g(x) + i\,h(x)$, where g and h are real-valued functions, can be differentiated in a natural way as

$$f'(x) = g'(x) + i\,h'(x).$$

6.2 Rules of differentiation

Constant multiple rule: If k is a constant and u is a function of x, then

$$(ku)' = ku', \quad \text{or equivalently} \quad \frac{d}{dx}(ku) = k\,\frac{du}{dx}.$$

Sum rule: If u and v are functions of x, then

$$(u+v)' = u' + v', \quad \text{or equivalently} \quad \frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}.$$

Product and **quotient rules**: If u and v are functions of x, then

$$(uv)' = u'v + uv', \quad \text{or equivalently} \quad \frac{d}{dx}(uv) = \frac{du}{dx}\,v + u\,\frac{dv}{dx},$$

and

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}, \quad \text{or equivalently} \quad \frac{d}{dx}\left(\frac{u}{v}\right) = \left(\frac{du}{dx}\,v - u\,\frac{dv}{dx}\right)\Big/ v^2.$$

This is sometimes called the **'function of a function' rule**.

Composite rule or **chain rule**: If g and u are two functions, and $h(x) = g(u(x))$, then

$$h'(x) = g'(u(x))\,u'(x).$$

Another way to write this is as

$$\frac{dh}{dx} = \frac{dh}{du}\,\frac{du}{dx}.$$

Implicit differentiation: Given an equation connecting two variables x and y, we can use implicit differentiation to calculate dy/dx by differentiating both sides with respect to x and then solving the resulting equation algebraically for dy/dx (instead of solving for y before differentiating). For example,

$$\frac{d}{dx}(x^3y^2) = 3x^2y^2 + 2x^3y\,\frac{dy}{dx}.$$

6.3 Standard derivatives

In each case, a is a constant.

Function	Derivative
a	0
x^a	ax^{a-1}
e^{ax}	ae^{ax}
$\ln(ax)$	$\dfrac{1}{x}$
$\sin(ax)$	$a\cos(ax)$
$\cos(ax)$	$-a\sin(ax)$
$\tan(ax)$	$a\sec^2(ax)$
$\cot(ax)$	$-a\operatorname{cosec}^2(ax)$
$\sec(ax)$	$a\sec(ax)\tan(ax)$
$\operatorname{cosec}(ax)$	$-a\operatorname{cosec}(ax)\cot(ax)$
$\arcsin(ax)$	$\dfrac{a}{\sqrt{1-a^2x^2}}$
$\arccos(ax)$	$-\dfrac{a}{\sqrt{1-a^2x^2}}$
$\arctan(ax)$	$\dfrac{a}{1+a^2x^2}$
$\operatorname{arccot}(ax)$	$-\dfrac{a}{1+a^2x^2}$
$\operatorname{arcsec}(ax)$	$\dfrac{a}{\lvert ax\rvert\sqrt{a^2x^2-1}}$
$\operatorname{arccosec}(ax)$	$-\dfrac{a}{\lvert ax\rvert\sqrt{a^2x^2-1}}$

6.4 Stationary points

The **gradient** of a function f at a point x_0 is the slope of the tangent to the graph of f at that point, and is given by the derivative of f at that point, i.e. $f'(x_0)$. A function is **increasing** on an interval if its gradient is positive throughout that interval; it is **decreasing** if its gradient is negative throughout that interval. A **stationary point** of f is a point x_0 where the gradient is zero, i.e. $f'(x_0) = 0$.

A function is **smooth** if it is continuous and has a continuous derivative.

Consider a stationary point x_0 of a smooth function f.

- x_0 is a **local maximum** if, for all x in the immediate vicinity of x_0, $f'(x) > 0$ if $x < x_0$ and $f'(x) < 0$ if $x > x_0$. An alternative condition is $f''(x_0) < 0$.
- x_0 is a **local minimum** if, for all x in the immediate vicinity of x_0, $f'(x) < 0$ if $x < x_0$ and $f'(x) > 0$ if $x > x_0$. An alternative condition is $f''(x_0) > 0$.
- x_0 is a **point of inflection** if $f''(x_0) = 0$ and $f''(x)$ changes sign as x increases through x_0.

A **global maximum** of a function f is a point x_0 such that $f(x_0) \geq f(x)$ for all x where f is defined. A **global minimum** of a function f is a point x_0 such that $f(x_0) \leq f(x)$ for all x where f is defined. A function f is **bounded above** by an **upper bound** A if $f(x) \leq A$ for all x where f is defined. A function f is **bounded below** by a **lower bound** B if $f(x) \geq B$ for all x where f is defined.

6.5 Curve sketching

The following is a possible procedure for sketching the graph of $y = f(x)$, where $f(x)$ is some given function.

1. Check whether $f(x)$ is a standard function whose graph you already know, or is a simple modification of such a function. If not, proceed to Step 2.
2. Determine how y behaves when x is very large and positive, and when x is very large and negative.
3. Look for any obvious symmetries or repetitions in the behaviour of f.
4. Find where the curve crosses the x- and y-axes, if at all.
5. Look for any values of x at which $f(x)$ is undefined, and examine the behaviour of $f(x)$ near these values of x.
6. Find the locations of any local maxima, local minima or points of inflection.
7. Try to determine whether there are any intervals over which the function is increasing or decreasing.

8. Transfer the information found in Steps 4 and 6 to a sketch graph, then use this information together with any information found in Steps 2, 3, 5 and 7 to try to sketch a smooth curve. If you are still unsure about any parts of the curve, choose suitable values of x and plot the corresponding points $(x, f(x))$ before completing the curve.

Lines $y = c$ where $f(x) \to c$ as $x \to \pm\infty$, and lines $x = c$ where $f(x) \to \pm\infty$ as $x \to c$, where c is a constant, are known as **asymptotes** of the graph of the function $f(x)$.

6.6 Taylor polynomials and series

Factorials

For any positive integer n, we define $\boldsymbol{n}$ **factorial**, written $n!$, by

$$n! = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n.$$

The first few factorials are $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$. We also define $0! = 1$.

Taylor polynomials

For a function $f(x)$ with n continuous derivatives near $x = a$, the **Taylor polynomial of degree $\boldsymbol{n}$** about $x = a$ or the **$\boldsymbol{n}$th-order Taylor polynomial** about $x = a$ is

$$p_n(x) = f(a) + (x-a)f'(a) + \frac{1}{2!}(x-a)^2 f''(a) + \cdots + \frac{1}{n!}(x-a)^n f^{(n)}(a).$$

When used to approximate $f(x)$ near $x = a$, we refer to this polynomial as the **$\boldsymbol{n}$th-order Taylor approximation** to $f(x)$ near $x = a$, and write

$$f(x) \simeq f(a) + (x-a)f'(a) + \frac{1}{2!}(x-a)^2 f''(a) + \cdots + \frac{1}{n!}(x-a)^n f^{(n)}(a).$$

In particular, $n = 1$ gives the **tangent approximation**

$$f(x) \simeq f(a) + (x-a)f'(a),$$

and $n = 2$ gives the **quadratic approximation**

$$f(x) \simeq f(a) + (x-a)f'(a) + \tfrac{1}{2}(x-a)^2 f''(a).$$

These approximations are good when x is close to a.

Taylor series

The **Taylor series about** $x = a$ for a function $f(x)$ with infinitely many continuous derivatives near $x = a$ is

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + \frac{1}{2!}(x-a)^2 f''(a) + \cdots \\ &\quad + \frac{1}{n!}(x-a)^n f^{(n)}(a) + \cdots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(x-a)^n f^{(n)}(a). \end{aligned}$$

Some standard Taylor series about $x = 0$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^{n-1}\frac{1}{(2n-1)!}x^{2n-1} + \cdots,$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^{n}\frac{1}{(2n)!}x^{2n} + \cdots,$$

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \cdots,$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n-1}\frac{1}{n}x^n + \cdots \quad (-1 < x < 1).$$

Small-angle approximations

If the angle θ is small (sometimes written $\theta \ll 1$) and is measured in radians, then we can obtain useful approximations by truncating the above Taylor series:

$$\sin\theta \simeq \theta,$$
$$\cos\theta \simeq 1 - \tfrac{1}{2}\theta^2.$$

7 Integration

7.1 Notation and terminology

The **indefinite integral** of a continuous function $f(x)$ is

$$\int f(x)\,dx = F(x) + C,$$

where F is a function such that $F'(x) = f(x)$, known as an **integral** or **antiderivative** of f, and C is a constant, often referred to as an **arbitrary constant** or **constant of integration**.

The **definite integral** of a continuous function $f(x)$ from a to b is

$$\int_a^b f(x)\,dx = \big[F(x)\big]_a^b = F(b) - F(a),$$

where F is any integral of f. The numbers a and b are called the **lower limit of integration** and **upper limit of integration**, respectively. If the areas bounded by the graph of $f(x)$ above and below the x-axis between a and b are A_1 and A_2, respectively, then

$$\int_a^b f(x)\,dx = A_1 - A_2.$$

The process of finding an indefinite or definite integral is known as **integration**, and the function f being integrated is known as the **integrand**. If F_1 and F_2 are two integrals of f, then they differ by a constant, i.e. $F_1(x) = F_2(x) + C$, where C is a constant.

7.2 Rules of integration

Constant multiple rule:

$$\int k\, f(x)\, dx = k \int f(x)\, dx \quad \text{(where } k \text{ is a constant)}.$$

Sum rule:

$$\int \big(f(x) + g(x)\big)\, dx = \int f(x)\, dx + \int g(x)\, dx.$$

Integration by substitution:

$$\int f\big(g(x)\big)\, g'(x)\, dx = \int f(u)\, du \quad \text{(where } u = g(x)\text{)},$$

or in Leibniz notation

$$\int f(u)\, \frac{du}{dx}\, dx = \int f(u)\, du.$$

The following formula, which can be derived by integration by substitution, is also useful:

$$\int \frac{g'(x)}{g(x)}\, dx = \ln|g(x)| + C \quad \text{(where } g(x) \neq 0\text{)}.$$

Integration by parts:

$$\int f(x)\, g'(x)\, dx = f(x)\, g(x) - \int f'(x)\, g(x)\, dx.$$

For definite integrals,

$$\int_a^b f(x)\, g'(x)\, dx = \big[f(x)\, g(x)\big]_a^b - \int_a^b f'(x)\, g(x)\, dx.$$

A function f is **odd** if $f(-x) = -f(x)$. If f is odd and a is a positive constant, then

$$\int_{-a}^{a} f(x)\, dx = 0.$$

7.3 Standard integrals

In each case, a is a non-zero constant, b is any constant, and n is any integer. When using the table below to obtain indefinite integrals, add an arbitrary constant.

Function	Integral
a	ax
$x^a \ (a \neq -1)$	$\dfrac{x^{a+1}}{a+1}$
$\dfrac{1}{ax+b}$	$\dfrac{1}{a}\ln\lvert ax+b\rvert$
e^{ax}	$\dfrac{1}{a}e^{ax}$
$\ln(ax)$	$x(\ln(ax)-1)$
$\sin(ax)$	$-\dfrac{1}{a}\cos(ax)$
$\cos(ax)$	$\dfrac{1}{a}\sin(ax)$
$\tan(ax)$	$-\dfrac{1}{a}\ln\lvert\cos(ax)\rvert$
$\cot(ax)$	$\dfrac{1}{a}\ln\lvert\sin(ax)\rvert$
$\sec(ax)$	$\dfrac{1}{a}\ln\lvert\sec(ax)+\tan(ax)\rvert$
$\operatorname{cosec}(ax)$	$\dfrac{1}{a}\ln\lvert\operatorname{cosec}(ax)-\cot(ax)\rvert$
$\sec^2(ax)$	$\dfrac{1}{a}\tan(ax)$
$\operatorname{cosec}^2(ax)$	$-\dfrac{1}{a}\cot(ax)$
$x\sin(ax)$	$\dfrac{1}{a^2}\big(\sin(ax)-ax\cos(ax)\big)$
$x\cos(ax)$	$\dfrac{1}{a^2}\big(\cos(ax)+ax\sin(ax)\big)$
$\dfrac{1}{x^2+a^2}$	$\dfrac{1}{a}\arctan\left(\dfrac{x}{a}\right)$
$\dfrac{1}{(x-a)(x-b)}$	$\dfrac{1}{a-b}\ln\left\lvert\dfrac{a-x}{x-b}\right\rvert$
$\dfrac{1}{\sqrt{x^2+a^2}}$	$\ln(x+\sqrt{x^2+a^2}\,)$
$\dfrac{1}{\sqrt{x^2-a^2}}$	$\ln\lvert x+\sqrt{x^2-a^2}\,\rvert$
$\dfrac{1}{\sqrt{a^2-x^2}}$	$\arcsin\left(\dfrac{x}{a}\right)$

Unit summaries

Unit 1 Getting started

The material in Unit 1 is summarised in Sections 1–7 of this Handbook.

Unit 2 First-order differential equations

1. A **differential equation** is an equation that relates an independent variable x, a dependent variable y, and one or more derivatives of y. Its **order** is the order of the highest derivative that appears. So a **first-order** differential equation for $y = y(x)$ contains its first derivative, dy/dx, but no higher derivative of y. This unit considers only first-order differential equations that can be expressed in the form

   $$\frac{dy}{dx} = f(x, y),$$

   where $f(x, y)$ stands for some expression that may contain either or both of the variables x and y.
2. A **solution** of a differential equation is a function $y = y(x)$ that satisfies it. A solution that is written in the form '$y =$ function of x' is an **explicit solution**; otherwise it is an **implicit solution**, i.e. an equation of the form $F(x, y) = 0$ for some function F.
3. The **general solution** of a differential equation is the collection of all of its solutions. It is usually possible to give the general solution of a first-order differential equation as a formula containing one **arbitrary constant**. A **particular solution** of a differential equation is a single solution containing no arbitrary constant.
4. An **initial condition** for a first-order differential equation $dy/dx = f(x, y)$ is an assignment of a value y_0 that the dependent variable y must take when the independent variable x takes some given value x_0. An initial condition may be specified in the form '$y = y_0$ when $x = x_0$' or '$y(x_0) = y_0$'; x_0 and y_0 are referred to as **initial values**.

   An **initial-value problem** is to find the particular solution of a differential equation that satisfies a given initial condition. Given the general solution of a differential equation involving an arbitrary constant C, we can determine the solution satisfying an initial condition by substituting the initial values into the general solution; this gives an equation from which the required value of C can, in principle, be found.

5. Some differential equations have an **analytic solution**, i.e. an explicit general solution derived using calculus. To decide whether an equation of the form $dy/dx = f(x, y)$ may be solved analytically by one of the methods described in the module, proceed as follows.

   (a) If $f(x, y)$ is independent of y, so that $f(x, y) = f(x)$, then the equation may be solved by **direct integration**: its general solution is $y = F(x) + C$, where $F(x)$ is an integral of $f(x)$ and C is an arbitrary constant.

   (b) If $f(x, y)$ has the form $f(x, y) = g(x)\,h(y)$, then use the method of separation of variables described below.

   (c) If $f(x, y)$ has the form $f(x, y) = h(x) - g(x)\,y$ (so that the equation is of the form $dy/dx + g(x)\,y = h(x)$), then the equation is linear, and may be solved by the integrating factor method described below.

6. To solve the differential equation $dy/dx = g(x)\,h(y)$, where $h(y) \neq 0$, using the method of **separation of variables**, proceed as follows.

   (a) Divide both sides of the equation by $h(y)$, and integrate with respect to x, to obtain

   $$\int \frac{1}{h(y)}\,dy = \int g(x)\,dx.$$

   (b) If possible, perform the two integrations, obtaining an implicit form of the general solution, which should include one arbitrary constant.

   (c) If possible, rearrange the formula found in Step (b) to give y in terms of x; this is the explicit general solution of the differential equation. In addition, there may be supplementary solutions if y is a constant satisfying $h(y) = 0$.

7. A first-order differential equation is **linear** if it can be expressed in the form

   $$\frac{dy}{dx} + g(x)\,y = h(x).$$

   It is **homogeneous** if $h(x) = 0$ for all x, **inhomogeneous** otherwise. For all the cases that you will meet in this module, the initial-value problem

   $$\frac{dy}{dx} + g(x)\,y = h(x), \quad y(x_0) = y_0,$$

   has a unique solution.

8. Linear first-order differential equations can be solved by the **integrating factor method**. An **integrating factor** for the equation $dy/dx + g(x)\,y = h(x)$ is a function $p(x)$ with the property that, after multiplication by $p(x)$, the left-hand side of the equation becomes an exact derivative:

$$p(x)\left(\frac{dy}{dx} + g(x)\,y\right) = \frac{d}{dx}\big(p(x)\,y\big).$$

Multiplication by an integrating factor therefore makes it possible to solve the equation by direct integration.

9. The integrating factor for the equation $dy/dx + g(x)\,y = h(x)$ is

$$p(x) = \exp\left(\int g(x)\,dx\right).$$

In the constant-coefficient case $dy/dx + Ay = h(x)$, where A is a constant, $p(x) = \exp(Ax)$.

The steps of the integrating factor method are as follows.

(a) Determine the integrating factor $p(x)$.

(b) Rewrite the equation as $\dfrac{d}{dx}\big(p(x)\,y\big) = p(x)\,h(x)$.

(c) Integrate to obtain $p(x)\,y = \displaystyle\int p(x)\,h(x)\,dx + C$, where C is a constant.

(d) Divide through by $p(x)$, to obtain the general solution in explicit form.

Unit 3 Second-order differential equations

1. A **second-order differential equation** is a differential equation that contains the second derivative d^2y/dx^2 of the dependent variable y with respect to the independent variable x, but no higher derivative. Thus in addition to d^2y/dx^2, any or all of the following may occur in the equation: x, y, dy/dx. The general solution of such an equation normally involves two arbitrary constants.

2. A **linear** second-order differential equation is one that can be written in the form

$$a(x)\,\frac{d^2y}{dx^2} + b(x)\,\frac{dy}{dx} + c(x)\,y = f(x),$$

where $a(x)$, $b(x)$, $c(x)$ and $f(x)$ are given functions. If $f(x)$ is identically zero, then the equation is **homogeneous**; otherwise it is **inhomogeneous**.

3. A **linear constant-coefficient** second-order differential equation is one in which the functions $a(x)$, $b(x)$ and $c(x)$ are constant (with $a \neq 0$), i.e.

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x),$$

and $f(x)$ is a given function.

4. The methods of solution described in the unit make use, explicitly or implicitly, of the following **principle of superposition**.

If $y_1(x)$ is a solution of $ay'' + by' + cy = f_1(x)$, and $y_2(x)$ is a solution of $ay'' + by' + cy = f_2(x)$, then for any constants k_1 and k_2,

$$y(x) = k_1\, y_1(x) + k_2\, y_2(x)$$

is a solution of the equation

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = k_1\, f_1(x) + k_2\, f_2(x).$$

5. The **general solution of the homogeneous equation**

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0$$

is obtained by following the procedure below.

(a) Solve for λ the **auxiliary equation**

$$a\lambda^2 + b\lambda + c = 0.$$

(b) (i) If the auxiliary equation has distinct real roots λ_1 and λ_2, then the general solution of the differential equation is

$$y(x) = Ce^{\lambda_1 x} + De^{\lambda_2 x}.$$

(ii) If the auxiliary equation has equal real roots $\lambda_1 = \lambda_2$ (i.e. $a\lambda^2 + b\lambda + c$ is a perfect square), then the general solution of the differential equation is

$$y(x) = (C + Dx)e^{\lambda_1 x}.$$

(iii) If the auxiliary equation has complex conjugate roots $\lambda_1 = \alpha + i\beta$ and $\lambda_2 = \alpha - i\beta$, then the general solution of the differential equation is

$$y(x) = e^{\alpha x}(C\cos\beta x + D\sin\beta x).$$

In each case, C and D are arbitrary constants.

6. For the oscillation represented by $x(t) = A\sin(\omega t + \phi)$, the constant $A \geq 0$ is called the **amplitude**, the constant ω is called the **angular frequency**, and the constant ϕ is called the **phase constant**. The **period** of the oscillation is $2\pi/\omega$.

7. The **general solution of the inhomogeneous equation**

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x)$$

is given by $y = y_c + y_p$, where:

- y_c, the **complementary function**, is the general solution of the **associated homogeneous equation**

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0$$

- y_p, a **particular integral**, is any particular solution of the original inhomogeneous equation.

This leads to the following procedure for finding the general solution of an inhomogeneous equation.

(a) Find the complementary function, by solving the auxiliary equation of the associated homogeneous equation.

(b) Find a particular integral, as described below.

(c) Add the particular integral to the complementary function.

8. For certain inhomogeneous equations, particular solutions can be found by the **method of undetermined coefficients**. This method works when the function $f(x)$ is a polynomial, an exponential function or a sinusoidal function, i.e. a linear combination of a sine and a cosine.

The method involves using a **trial solution** $y(x)$ of a form similar to that of $f(x)$, but with coefficients that initially are undetermined. The coefficients are determined by substituting the trial solution into the differential equation and choosing the coefficients so that the equation is satisfied.

Suitable trial solutions $y(x)$ for specified right-hand-side (or target) functions $f(x)$ are shown in the following table.

Target function $f(x)$	Trial solution $y(x)$
$m_n x^n + m_{n-1}x^{n-1} + \cdots + m_1 x + m_0$	$p_n x^n + p_{n-1}x^{n-1} + \cdots + p_1 x + p_0$
me^{kx}	pe^{kx}
$m\cos kx + n\sin kx$	$p\cos kx + q\sin kx$

The coefficients in the left-hand column, $m, k, m_0, m_1, \ldots$, are given constants. Those in the right-hand column, $p, p_0, p_1, \ldots$, are constants to be determined, by differentiating the expression for $y(x)$ twice, substituting into the left-hand side of the differential equation, and equating coefficients of corresponding terms on the left- and right-hand sides.

Note that even if some of the given coefficients in $f(x)$ are zero – for example $f(x) = 3x$ or $f(x) = 2\cos x$ – it is still necessary in general to use the full expression from the right-hand column of the table.

9. An **initial-value problem** for a second-order differential equation is a problem in which one has to find the particular solution $y = y(x)$ of a given equation such that y and its derivative y' take specified values y_0 and z_0, respectively, when the independent variable x takes the value x_0. The numbers x_0, y_0 and z_0 are called **initial values**. The relationships between initial values are called **initial conditions**. These may be specified either as '$y = y_0$ and $y' = z_0$ when $x = x_0$', or as '$y(x_0) = y_0$, $y'(x_0) = z_0$'.

   An initial-value problem has a unique solution, which can be obtained by finding the general solution of the differential equation and substituting in the initial values to determine the two arbitrary constants that it contains. It is necessary first to differentiate the general solution to apply the initial condition $y'(x_0) = z_0$.

10. In a **boundary-value problem**, a condition is placed on the value of either y or its derivative, or some combination of the two, at each of two different values of x. The conditions are referred to as **boundary conditions**, and values of x and y in these conditions are **boundary values**. Such a problem may have a unique solution, or no solution, or an infinite number of solutions.

Unit 4 Vectors and matrices

1. A **vector** is a mathematical object consisting of a non-negative real number called its **magnitude**, and a **direction**.

   In printed text, vectors appear in **bold**; in handwritten work they are written underlined.

   - Two vectors are **equal** if and only if they have the same magnitude and direction.
   - The magnitude of a vector $\mathbf{v}$ is denoted by $|\mathbf{v}|$ or sometimes by v.
   - The **zero vector 0** has magnitude zero; no direction is defined for it.
   - The **displacement** from a point P to a point Q is represented by the **displacement vector** $\overrightarrow{PQ}$.

2. In discussions involving vectors, the word **scalar** is used to denote a real number (positive, negative or zero).

3. **Scaling a vector** or **scalar multiplication** of a vector is the process of multiplying a vector by a scalar.

   For a vector $\mathbf{v}$ and (non-zero) scalar m, the **scalar multiple** $m\mathbf{v}$ is the vector whose magnitude is $|m|\,|\mathbf{v}|$, and whose direction is

   - in the same direction as $\mathbf{v}$ if $m > 0$
   - in the opposite direction to $\mathbf{v}$ if $m < 0$.

Note that $-\mathbf{v} = (-1)\mathbf{v}$ is the vector with the same magnitude as $\mathbf{v}$ but pointing in the opposite direction (such vectors are called **antiparallel**).

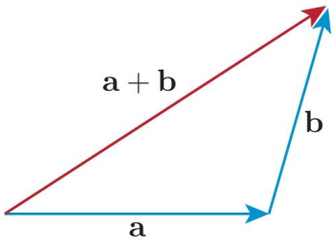
Also note that $0\mathbf{v} = m\mathbf{0} = \mathbf{0}$ for any vector $\mathbf{v}$ and scalar m.

4. A **unit vector** is a vector whose magnitude is 1. For $\mathbf{v} \neq \mathbf{0}$, the vector $\widehat{\mathbf{v}} = (1/|\mathbf{v}|)\mathbf{v}$ is a unit vector in the same direction as $\mathbf{v}$. Unit vectors along the Cartesian axes are called **Cartesian unit vectors**, and are denoted by $\mathbf{i}$, $\mathbf{j}$ and (in three dimensions) $\mathbf{k}$.

5. **Vector addition** is defined geometrically by the **triangle rule** (see the diagram in the margin).

   The vector $\mathbf{a} + \mathbf{b}$ is called the **sum** or **resultant** of $\mathbf{a}$ and $\mathbf{b}$. **Vector subtraction** is defined by $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$.

   Note that $\mathbf{a} - \mathbf{a} = \mathbf{0}$ and $\mathbf{0} + \mathbf{a} = \mathbf{a}$ for any vector $\mathbf{a}$.



Triangle rule

6. A three-dimensional Cartesian coordinate system consists of three mutually perpendicular axes, usually labelled x, y and z, that meet at a point called the origin.

   A **right-handed coordinate system** is one satisfying the **right-hand rule**:

   - Point the straightened fingers of your right hand in the direction of the positive x-axis, and rotate your wrist until you find that you can bend your fingers in the direction of the positive y-axis.
   - Extend the thumb of your right hand. This is the direction of the positive z-axis for a right-handed coordinate system.

7. Given a Cartesian coordinate system, any vector $\mathbf{a}$ can be written uniquely in **component form** as

   $$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}, \quad \text{or equivalently} \quad \mathbf{a} = (a_x, a_y, a_z),$$

   where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors in the directions of the positive x-, y- and z-axes, respectively. The numbers a_x, a_y and a_z are called the (**Cartesian**) **components** of $\mathbf{a}$ in the directions of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. The process of finding the Cartesian components of a vector is called **resolving a vector into its components**.

   Let $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$. Then:

   - $|\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$
   - $\mathbf{a} + \mathbf{b} = (a_x + b_x)\mathbf{i} + (a_y + b_y)\mathbf{j} + (a_z + b_z)\mathbf{k}$
   - $m\mathbf{a} = (ma_x)\mathbf{i} + (ma_y)\mathbf{j} + (ma_z)\mathbf{k}$.

Here and below, we give the formulas for the three-dimensional case; the corresponding two-dimensional versions are obtained by omitting the third component.

8. The **position vector** of a point A is the displacement vector $\overrightarrow{OA}$, where O is the origin. So if A has coordinates (x, y, z), then its position vector is $\overrightarrow{OA} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

   If P and Q are two points with coordinates (x_P, y_P, z_P) and (x_Q, y_Q, z_Q), respectively, then the displacement vector $\overrightarrow{PQ}$ is

   $$\overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = (x_Q - x_P)\mathbf{i} + (y_Q - y_P)\mathbf{j} + (z_Q - z_P)\mathbf{k}.$$

   The **vector equation of the straight line** joining points P and Q whose position vectors are $\mathbf{p}$ and $\mathbf{q}$, respectively, is

   $$\mathbf{r}(t) = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) = (1 - t)\mathbf{p} + t\mathbf{q}.$$

   If t is allowed to vary over all real numbers, then $\mathbf{r}(t)$ traces out the entire straight line through the points P and Q. If $0 \leq t \leq 1$, then $\mathbf{r}(t)$ traces out the line segment from P to Q.

9. The **scalar product** (or **dot product**) of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the scalar defined by

   $$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \cos\theta,$$

   where θ (for $0 \leq \theta \leq \pi$) is the angle between the directions of $\mathbf{a}$ and $\mathbf{b}$. In particular, non-zero vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular if and only if $\mathbf{a} \cdot \mathbf{b} = 0$. Also, $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$.

   - The scalar products of the Cartesian unit vectors are

     $$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1, \quad \mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0.$$

   - In component form, if $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$, then

     $$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z.$$

     The formula $|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a} = a_x^2 + a_y^2 + a_z^2$ is a particular case.

   - The angle θ between two (non-zero) vectors $\mathbf{a}$ and $\mathbf{b}$ is given by

     $$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\sqrt{b_x^2 + b_y^2 + b_z^2}}.$$

10. The component of a vector $\mathbf{a}$ in the direction of a unit vector $\widehat{\mathbf{u}}$ is

    $$a_u = \mathbf{a} \cdot \widehat{\mathbf{u}} = |\mathbf{a}| \cos\theta,$$

    where θ is the angle between $\mathbf{a}$ and $\widehat{\mathbf{u}}$.

    The Cartesian components of $\mathbf{a}$ are $\mathbf{a} \cdot \mathbf{i}$, $\mathbf{a} \cdot \mathbf{j}$ and $\mathbf{a} \cdot \mathbf{k}$; in other words,

    $$\mathbf{a} = (\mathbf{a} \cdot \mathbf{i})\,\mathbf{i} + (\mathbf{a} \cdot \mathbf{j})\,\mathbf{j} + (\mathbf{a} \cdot \mathbf{k})\,\mathbf{k}.$$

11. The **vector product** (or **cross product**) of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the vector defined by

$$\mathbf{a} \times \mathbf{b} = (|\mathbf{a}|\,|\mathbf{b}| \sin\theta)\,\widehat{\mathbf{n}},$$

where θ (for $0 \leq \theta \leq \pi$) is the angle between the directions of $\mathbf{a}$ and $\mathbf{b}$, and $\widehat{\mathbf{n}}$ is a unit vector at right angles to both $\mathbf{a}$ and $\mathbf{b}$, whose sense is given by the right-hand rule for vector products: this is like the right-hand rule for a coordinate system given above, but with the x-, y-, z-axes replaced by $\mathbf{a}$, $\mathbf{b}$, $\widehat{\mathbf{n}}$, respectively.

The vector product has the following algebraic properties:

- $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$.
- $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$ and $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c})$.
- For any scalar λ, $(\lambda\mathbf{a}) \times \mathbf{b} = \lambda(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (\lambda\mathbf{b})$.
- $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ if and only if one of $\mathbf{a}$ and $\mathbf{b}$ is a scalar multiple of the other – in other words, one of $\mathbf{a}$ and $\mathbf{b}$ is zero, or they are parallel or antiparallel. (In particular, $\mathbf{a} \times \mathbf{a} = \mathbf{0}$.)
- In general, $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$.
- The vector products of the Cartesian unit vectors are

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0},$$
$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}.$$

- In component form, the vector product of the vectors $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ and $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$ is

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y)\,\mathbf{i} + (a_z b_x - a_x b_z)\,\mathbf{j} + (a_x b_y - a_y b_x)\,\mathbf{k}.$$

There is an easy way to remember this formula in terms of determinants: see items 23 and 27.

12. The **area of a parallelogram** whose sides are defined by vectors $\mathbf{a}$ and $\mathbf{b}$ is $|\mathbf{a} \times \mathbf{b}|$.

The **area of a triangle** two of whose sides are $\mathbf{a}$ and $\mathbf{b}$ is $\frac{1}{2}|\mathbf{a} \times \mathbf{b}|$.

The **volume of a parallelepiped** whose sides through one vertex are $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, is $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$.

13. A **matrix** is a rectangular array of **elements** (usually numbers) arranged in rows and columns. If a matrix has m rows and n columns, then it is said to be an $m \times n$ matrix (read as 'm by n matrix'), or to be of **order** $m \times n$.

- An $m \times n$ matrix $\mathbf{A}$ that has a_{ij} as its element in the ith row and the jth column is written as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

or $\mathbf{A} = [a_{ij}]$ for short.

- A matrix with one column is often referred to as a **column vector**; a matrix with one row is a **row vector**. For example, a vector such as $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ may be written as the column vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{or} \quad \mathbf{a} = [a_1 \quad a_2 \quad a_3]^T.$$

The second form here is simply an alternative to the first, used to save space. (The superscript T stands for 'transpose', which is discussed in item 19.)

- A **square matrix** is one that has the same number of rows as columns.
- The $m \times n$ **zero matrix 0** is the $m \times n$ matrix all of whose elements are zero.
- Two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ are **equal** if they both have order $m \times n$, and $a_{ij} = b_{ij}$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

14. If $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ have the same order, then their **sum** is a matrix of the same order, given by $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$. Thus the elements of $\mathbf{A} + \mathbf{B}$ are obtained by adding together the elements of $\mathbf{A}$ and $\mathbf{B}$ in the same positions. For example, for column vectors we have

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{bmatrix}.$$

Similarly for subtraction.

15. The **scalar multiple** of a matrix $\mathbf{A} = [a_{ij}]$ by a number λ is given by

$$\lambda\mathbf{A} = [\lambda a_{ij}].$$

Thus $\lambda\mathbf{A}$ is obtained by multiplying each of the elements of $\mathbf{A}$ by λ. For example, for column vectors we have

$$\lambda \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \lambda a_1 \\ \lambda a_2 \\ \lambda a_3 \end{bmatrix}.$$

16. Matrix addition and scalar multiplication obey obvious rules of algebra. For any matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of the same order, and any scalar λ, the algebra is:

- **commutative**, $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- **associative**, $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- **distributive**, $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$.

17. The **product** of an $m \times p$ matrix $\mathbf{A} = [a_{ij}]$ and a $p \times n$ matrix $\mathbf{B} = [b_{ij}]$ is the $m \times n$ matrix $\mathbf{C} = \mathbf{AB}$, where $\mathbf{C} = [c_{ij}]$ is formed using the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$, to give

$$c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj}.$$

We imagine placing the ith row of $\mathbf{A}$ on the jth column of $\mathbf{B}$, multiplying the pairs of elements together, and then adding the products.

The product of two matrices can be formed only if the number of columns of the first matrix is the same as the number of rows of the second. For example, the product of the matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the column vector $\mathbf{r} = [x \quad y]^T$ is given by

$$\mathbf{Ar} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

18. In general, matrix multiplication is not commutative, so $\mathbf{AB}$ may not be equal to $\mathbf{BA}$.

For any matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, of the appropriate sizes so that all products can be formed, matrix multiplication is:

- associative, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- distributive over addition, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

The **powers** of a square matrix are defined in the obvious way: $\mathbf{A}^2 = \mathbf{AA}$, $\mathbf{A}^3 = \mathbf{AAA}$, and so on.

19. The **transpose** $\mathbf{A}^T$ of a matrix $\mathbf{A}$ is the matrix obtained by interchanging the rows and columns of $\mathbf{A}$. If we denote the matrix $\mathbf{A}$ by $[a_{ij}]$ and $\mathbf{A}^T$ by $[a_{ij}^T]$, then $a_{ij}^T = a_{ji}$. If $\mathbf{A}$ is an $m \times n$ matrix, then $\mathbf{A}^T$ is an $n \times m$ matrix.

- For any matrices $\mathbf{A}$ and $\mathbf{B}$ of the same order, $(\mathbf{A}^T)^T = \mathbf{A}$ and $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
- For any two matrices that can be multiplied,

  $$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T;$$

  note the change in order.

20. The **identity matrix** is a square matrix $\mathbf{I}$ whose elements are all zero except the diagonal elements, which equal 1. For example, the 2×2 identity matrix is $\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Assuming that all the matrix products make sense, the identity matrix satisfies $\mathbf{AI} = \mathbf{A}$ and $\mathbf{IA} = \mathbf{A}$.

21. The **inverse** of a square matrix $\mathbf{A}$ is the matrix, usually denoted $\mathbf{A}^{-1}$, of the same order such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

    Only a square matrix can have an inverse, but many square matrices do not have inverses; those that do not are called **non-invertible** (or **singular**), while those that do are called **invertible** (or **non-singular**).

    - $\mathbf{I}^{-1} = \mathbf{I}$.
    - If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices of the same size, then $\mathbf{AB}$ is invertible and
      $$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1};$$
      note the change in order.
    - A 2×2 matrix
      $$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
      is invertible if and only if $ad - bc \neq 0$. When this condition holds, we can evaluate the inverse using the formula
      $$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

22. If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then its **determinant** is a scalar, denoted by $\det \mathbf{A}$ or $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$, given by $\det \mathbf{A} = ad - bc$. Thus $\mathbf{A}$ is invertible (non-singular) if $\det \mathbf{A} \neq 0$, and non-invertible (singular) if $\det \mathbf{A} = 0$.

23. The determinant of the 3×3 matrix
    $$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$
    is
    $$\begin{aligned} \det \mathbf{A} &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1). \end{aligned} \tag{3}$$

24. The determinant of an $n \times n$ matrix for $n \geq 2$ can be evaluated using **Laplace's rule**.

    Given an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, its determinant $\det \mathbf{A}$ may be expanded in terms of the elements in row i and their cofactors C_{ij} as
    $$\det \mathbf{A} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in},$$
    where $C_{ij} = (-1)^{i+j}M_{ij}$, and M_{ij} is the minor obtained by deleting row i and column j of the original determinant and forming the determinant of what remains.

    This rule can be used to derive equation (3).

Although we are free to use any row or column when applying Laplace's rule, it is usual to use the first row.

25. All determinants have the following properties.
    - Interchanging any two rows or any two columns of $\mathbf{A}$ changes the sign of $\det \mathbf{A}$.
    - $\det(\mathbf{A}^T) = \det \mathbf{A}$.
    - Multiplying any row or any column of $\mathbf{A}$ by a scalar k multiplies $\det \mathbf{A}$ by k.
    - For any scalar k and any $n \times n$ matrix $\mathbf{A}$, $\det(k\mathbf{A}) = k^n \det \mathbf{A}$.
    - Adding a multiple of one row of $\mathbf{A}$ to another row does not change $\det \mathbf{A}$. Likewise for columns.
    - If a matrix $\mathbf{A}$ is invertible, then $\det(\mathbf{A}) \neq 0$ and $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.
    - For any two matrices that can be multiplied, $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$.
26. To find the inverse of an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, do the following.
    - Evaluate $\det \mathbf{A}$ and confirm that $\det \mathbf{A} \neq 0$. (If $\det \mathbf{A} = 0$, then no inverse exists.)
    - Evaluate the cofactor C_{ij} of each element a_{ij}, using the relation $C_{ij} = (-1)^{i+j} M_{ij}$, where M_{ij} is the minor obtained by deleting row i and column j of the original determinant and forming the determinant of what remains.
    - Form the square matrix $\mathbf{C} = [C_{ij}]$ composed of the cofactors C_{ij}.
    - Take the transpose of $\mathbf{C}$ to obtain the matrix $\mathbf{C}^T$.
    - Scale the matrix $\mathbf{C}^T$ by $1/\det \mathbf{A}$ to obtain the inverse of $\mathbf{A}$:
    $$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \mathbf{C}^T.$$
27. The vector product $\mathbf{b} \times \mathbf{c}$ of vectors $\mathbf{b} = [b_x \quad b_y \quad b_z]^T$ and $\mathbf{c} = [c_x \quad c_y \quad c_z]^T$ can be expressed as a 3×3 determinant:
    $$\mathbf{b} \times \mathbf{c} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}.$$
28. Let $\mathbf{a} = [a_x \quad a_y \quad a_z]^T$, $\mathbf{b} = [b_x \quad b_y \quad b_z]^T$ and $\mathbf{c} = [c_x \quad c_y \quad c_z]^T$. Then the **scalar triple product** $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ is given by
    $$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}.$$
    - The volume of the parallelepiped with sides defined by the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is given by the modulus of the scalar triple product, and therefore by the modulus of the determinant above.
    - Because interchanging rows of a matrix just changes the sign of its determinant, the scalar triple product obeys the **cyclic identity**
    $$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}).$$

29. Any **linear transformation** of the plane can be represented by a 2×2 matrix $\mathbf{A}$, called the **transformation matrix**. Its effect on the position vector $\mathbf{r}$ is to transform it into the matrix product $\mathbf{Ar}$. So if $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\mathbf{r} = [x \quad y]^T$, then

$$\mathbf{Ar} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

So the matrix $\mathbf{A}$ transforms the position vector $\mathbf{r} = [x \quad y]^T$ to the position vector $[ax + by \quad cx + dy]^T$.

- The Cartesian unit vectors $\mathbf{i} = [1 \quad 0]^T$ and $\mathbf{j} = [0 \quad 1]^T$ are transformed into the columns of the transformation matrix $\mathbf{A}$.
- The unit square whose sides are $\mathbf{i}$ and $\mathbf{j}$ is transformed into a parallelogram with area $|\det \mathbf{A}|$, the modulus of $\det \mathbf{A}$.

  A singular (non-invertible) matrix transforms the unit square into a line (or a point if $\mathbf{A} = \mathbf{0}$).
- Performing two successive transformations, represented by first $\mathbf{A}$ then $\mathbf{B}$, is equivalent to the effect of a single transformation, represented by the matrix product $\mathbf{BA}$ (note the order).
- The inverse $\mathbf{A}^{-1}$ of a transformation matrix $\mathbf{A}$ reverses the effect of the transformation.
- The **dilation matrix**

  $$\mathbf{D}(\kappa, \lambda) = \begin{bmatrix} \kappa & 0 \\ 0 & \lambda \end{bmatrix}$$

  is a linear transformation that rescales the plane by κ in the x-direction and λ in the y-direction.

  The inverse of a dilation matrix is another dilation matrix: $\mathbf{D}^{-1}(\kappa, \lambda) = \mathbf{D}(1/\kappa, 1/\lambda)$. This exists only if $\kappa \neq 0$ and $\lambda \neq 0$.
- The **rotation matrix**

  $$\mathbf{R}(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

  is a linear transformation that rotates the plane by an angle α about the origin, in the anticlockwise direction.

  The inverse of the rotation matrix $\mathbf{R}(\alpha)$ is $\mathbf{R}^{-1}(\alpha) = \mathbf{R}(-\alpha)$.

Unit 5 Linear algebra

1. An equation involving n variables $x_1, x_2, \ldots, x_n$ is said to be **linear** in each of those variables if it can be written in the form

   $$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = b,$$

   where $a_1, a_2, \ldots, a_n$ and b are constants.

2. Consider a **system of linear equations**

   $$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n, \end{aligned}$$

   where there is the same number n of equations as of unknowns $x_1, x_2, \ldots, x_n$. Such a system can be written in matrix form as

   $$\mathbf{Ax} = \mathbf{b}.$$

   Here $\mathbf{A}$ is the $n \times n$ square matrix $\mathbf{A} = [a_{ij}]$, called the **coefficient matrix**, formed from the coefficients a_{ij}. The vectors $\mathbf{x}$ and $\mathbf{b}$ are the column vectors $\mathbf{x} = [x_1 \quad x_2 \quad \ldots \quad x_n]^T$ and $\mathbf{b} = [b_1 \quad b_2 \quad \ldots \quad b_n]^T$.

   Usually the elements of $\mathbf{A}$ and $\mathbf{b}$ are given, while those of $\mathbf{x}$ are unknown and to be solved for.

   The **augmented matrix** of the system is the $n \times (n + 1)$ matrix that has the coefficient matrix for its first n columns and whose final column is the column vector $\mathbf{b}$ consisting of the right-hand sides of the equations. The augmented matrix is usually written $\mathbf{A}|\mathbf{b}$, with a vertical bar separating the coefficient matrix $\mathbf{A}$ from the right-hand-side vector $\mathbf{b}$.

3. The **leading** or **main diagonal** of a matrix $\mathbf{A} = [a_{ij}]$ is the collection of elements a_{ii}. For a square matrix it is the diagonal that runs from the top left corner to the bottom right corner.

4. An **upper triangular matrix** is a square matrix in which each entry below the leading diagonal is 0. A **lower triangular matrix** is a square matrix in which each entry above the leading diagonal is 0. A matrix that is upper triangular, lower triangular or both (i.e. **diagonal**) is sometimes referred to simply as a **triangular matrix**.

5. In principle, any system of equations $\mathbf{Ax} = \mathbf{b}$, for $\mathbf{A}$ non-singular, can be solved by calculating the inverse of $\mathbf{A}$ using the methods of Unit 4, then setting $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. However, this method is increasingly inefficient for systems of three equations or more. A more efficient method is to use Gaussian elimination.

6. **Gaussian elimination** is an efficient and systematic method of obtaining the solution $\mathbf{x}$ of a system of linear equations $\mathbf{Ax} = \mathbf{b}$.

   To solve a system of n linear equations in n unknowns, with coefficient matrix $\mathbf{A}$ and right-hand-side vector $\mathbf{b}$, by Gaussian elimination, carry out the following steps (if possible).

   (a) *Formulation*: Write down the augmented matrix $\mathbf{A}|\mathbf{b}$, denoting its rows by $\mathbf{R}_1, \ldots, \mathbf{R}_n$.

   (b) *Elimination*: Adapt the following row operations as necessary.

   (i) Subtract a multiple of $\mathbf{R}_1$ from $\mathbf{R}_2$, to reduce to zero the first element in the first column below the leading diagonal.

   (ii) Similarly, subtract a multiple of $\mathbf{R}_1$ from $\mathbf{R}_3, \ldots, \mathbf{R}_n$ to reduce to zero all the other elements in the first column below the leading diagonal.

   (iii) In the new matrix obtained, subtract multiples of $\mathbf{R}_2$ from $\mathbf{R}_3, \ldots, \mathbf{R}_n$ to reduce to zero all the elements in the second column below the leading diagonal.

   (iv) Continue this process until $\mathbf{A}|\mathbf{b}$ is reduced to $\mathbf{U}|\mathbf{c}$, where $\mathbf{U}$ is an upper triangular matrix.

   (c) *Solution*: Solve the system of equations with coefficient matrix $\mathbf{U}$ and right-hand-side vector $\mathbf{c}$ by **back substitution** – that is, solve the nth row for x_n; use this solution to solve the $(n-1)$th row for x_{n-1}; continue in this way, working up the rows, at each stage substituting values already known.

Performing a row operation really involves three sub-steps:
- write down the plan
- implement the plan
- relabel the rows.

7. Sometimes during the Gaussian elimination procedure, an unexpected zero can turn up on the leading diagonal element of a row. This can make it impossible to use the row for the elimination procedure. To remedy this, interchange the row with one below, then continue as normal.

8. The system of linear equations represented by the matrix equation $\mathbf{Ax} = \mathbf{b}$ is said to be **singular** when the coefficient matrix $\mathbf{A}$ is singular (i.e. when $\det \mathbf{A} = 0$). Such a system does not possess a unique solution but, depending on details, may have no solution or an infinity of solutions. When the system has no solution, it is sometimes said to be **inconsistent**.

9. If Gaussian elimination produces an upper triangular coefficient matrix $\mathbf{U}$ in which the final element on the leading diagonal is zero, then the back substitution process breaks down and the system is singular. In this case, if the final element in the final column is non-zero, then the system has no solutions, and if it is zero, then there may be an infinite number of solutions.

   The equations have a unique solution if and only if none of the elements on the leading diagonal of $\mathbf{U}$ is zero.

10. The n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **linearly dependent** if there are numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$, not all zero, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}.$$

In this case we can express one of the vectors as a linear combination of the others.

The vectors are **linearly independent** if the only solution of the equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$$

is $\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_n = 0$.

Two (non-zero) vectors are linearly independent if they are not collinear, i.e. parallel or antiparallel. Three (non-zero) vectors are linearly independent if they are not coplanar, i.e. if they do not all lie in the same plane.

11. The space to which vectors belong is often called a **vector space**. The **dimension** of a vector space is equal to the maximum number of linearly independent vectors that it allows.

    In an n-dimensional vector space, if you have n linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then you can express any other vector as a linear combination of them:

    $$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

    The set of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is called a **basis** for the n-dimensional vector space.

12. An **eigenvector** of a square matrix $\mathbf{A}$ is a non-zero vector $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ. The number λ is the **eigenvalue** corresponding to the eigenvector $\mathbf{v}$. Any non-zero scalar multiple of an eigenvector is also an eigenvector with the same eigenvalue.

    - Eigenvalues are distinct if they have different values; otherwise, they are said to be repeated. Eigenvectors are said to be distinct if they are linearly independent.
    - If a matrix has n distinct eigenvalues, then the corresponding n eigenvectors are all linearly independent.
    - If some of the eigenvalues of a matrix are the same, then the corresponding eigenvectors may or may not be linearly independent; further investigation is required.

13. If an $n \times n$ matrix has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then these can be used as a basis for n-dimensional vectors.

    So any n-dimensional vector $\mathbf{v}$ can be written as

    $$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

    for some scalars $c_1, c_2, \ldots, c_n$. This is called the **eigenvector expansion** of $\mathbf{v}$.

    The eigenvector expansion can be used to show that for (almost) any vector $\mathbf{v}$ and for large k, $\mathbf{A}^k\mathbf{v}$ is proportional to the eigenvector of $\mathbf{A}$ that has the eigenvalue with the largest modulus.

14. An $n \times n$ matrix $\mathbf{A}$ has **characteristic equation**

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0,$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. The left-hand side of this equation is a polynomial of degree n in λ. Its roots are the eigenvalues of $\mathbf{A}$.

15. In principle, the eigenvalues and eigenvectors of a square matrix $\mathbf{A}$ can be found as follows. Find the roots of the characteristic equation of $\mathbf{A}$ to obtain the eigenvalues. For each eigenvalue λ, solve the **eigenvector equations** $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ to obtain the corresponding eigenvector.

    If the eigenvalue is not repeated, then this will determine the eigenvector up to a scalar multiple. It is usually sufficient to choose some value for the scalar multiple to determine a suitable representative eigenvector.

16. To find the eigenvalues of the 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, do the following.

    (a) Write down the characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

    (b) Expand this as

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = \lambda^2 - (a + d)\lambda + (ad - bc) = 0.$$

    (c) Solve this quadratic equation to find the two values of λ, which are the required eigenvalues.

    For the above 2×2 matrix $\mathbf{A}$, the eigenvalues are

$$\lambda = \tfrac{1}{2}\left(a + d \pm \sqrt{(a - d)^2 + 4bc}\right).$$

    To find the eigenvector corresponding to the eigenvalue λ, do the following.

    (a) Write down the eigenvector equations

$$\begin{aligned} (a - \lambda)x + by &= 0, \\ cx + (d - \lambda)y &= 0. \end{aligned}$$

    (b) This pair of equations typically reduces to a single equation that is readily solved for x and y. The eigenvector is given by $\mathbf{v} = [x \quad y]^T$, with x and y replaced by their solved values. Any non-zero scalar multiple is also an eigenvector.

17. The **trace** of a square matrix is the sum of the elements on its leading diagonal. The trace of $\mathbf{A}$ is denoted by $\operatorname{tr}\mathbf{A}$.

    The characteristic equation of a 2×2 matrix can be written as

$$\lambda^2 - \operatorname{tr}\mathbf{A}\,\lambda + \det\mathbf{A} = 0.$$

18. If the elements of the above 2×2 matrix $\mathbf{A}$ are real, then the value of the discriminant $D = (a-d)^2 + 4bc$ characterises its eigenvalues:
    - If $D > 0$, then there are two distinct real eigenvalues.
    - If $D < 0$, then the two eigenvalues are complex conjugates of each other.
    - If $D = 0$, then the two eigenvalues are identical.
19. There are a number of general rules that apply to the eigenvalues and eigenvectors of an $n \times n$ matrix $\mathbf{A}$. To understand these rules, we need some definitions about matrices:
    - A **real matrix** is one whose elements are all real.
    - A **symmetric matrix** is one that is equal to its own transpose: $\mathbf{A} = \mathbf{A}^T$.

    Eigenvalue rules
    - The product of the eigenvalues of $\mathbf{A}$ is $\det \mathbf{A}$.
    - The sum of the eigenvalues of $\mathbf{A}$ is $\operatorname{tr} \mathbf{A}$.
    - Any complex eigenvalues of a real matrix occur in complex conjugate pairs, i.e. if λ is a complex eigenvalue, then so is $\overline{\lambda}$.
    - The eigenvalues of a triangular matrix are the diagonal entries.
    - The eigenvalues of a real symmetric matrix are real.
    - A matrix is non-invertible if and only if at least one of its eigenvalues is 0.

    Eigenvector rules
    - If a matrix has n distinct eigenvalues, then the corresponding n eigenvectors are all linearly independent.
    - For a real matrix, the components of an eigenvector corresponding to a real eigenvalue can be chosen to be real.
    - For a real matrix, if λ is a complex eigenvalue with corresponding eigenvector $\mathbf{v}$, then $\overline{\mathbf{v}}$, the vector whose components are the complex conjugates of those of $\mathbf{v}$, is an eigenvector corresponding to the eigenvalue $\overline{\lambda}$.
    - For any real symmetric matrix $\mathbf{A}$, the eigenvalues are real, and the eigenvectors $\mathbf{v}_i$ may be chosen to be real and orthogonal to each other, i.e. $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$.

Using matrices, the inner product $\mathbf{v}_i \cdot \mathbf{v}_j$ is calculated using $\mathbf{v}_i^T \mathbf{v}_j$.

Unit 6 Systems of linear differential equations

1. A **system of linear constant-coefficient first-order differential equations** for unknowns $x_1, x_2, \ldots, x_n$ takes the form

$$\begin{aligned}
\dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + h_1(t),\\
\dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + h_2(t),\\
&\vdots\\
\dot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + h_n(t),
\end{aligned}$$

where there is the same number n of equations as of unknowns.

The independent variable t often represents time. The dependent variables are $x_1(t), x_2(t), \ldots, x_n(t)$; but for systems with $n = 2$ or $n = 3$, x, y and z are often used instead of x_1, x_2 and x_3. The coefficients a_{ij} are constants.

The equations may be expressed in matrix form as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h},$$

where

$$\mathbf{x} = [x_1 \quad x_2 \quad \ldots \quad x_n]^T \quad \text{and} \quad \dot{\mathbf{x}} = [\dot{x}_1 \quad \dot{x}_2 \quad \ldots \quad \dot{x}_n]^T;$$

here $\mathbf{A} = [a_{ij}]$ is a constant matrix called the **matrix of coefficients**, but the vector $\mathbf{h} = [h_1 \quad h_2 \quad \ldots \quad h_n]^T$ may depend on t.

Note that there are exceptions to this rule, but we do not consider such cases in this module.

2. The **general solution** of a first-order system of n equations can be expressed as a solution containing n arbitrary constants.

A **particular solution** is a solution containing no arbitrary constants and satisfying given conditions.

Particular solutions of systems of n first-order differential equations are usually obtained by demanding that the general solution satisfies n initial conditions. These conditions fix the n arbitrary constants.

3. A system with $\mathbf{h} = \mathbf{0}$, i.e. of the form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, is said to be **homogeneous**. An **inhomogeneous** system is one where $\mathbf{h} \neq \mathbf{0}$.

To find the general solution of $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an $n \times n$ matrix, proceed as follows.

If the matrix $\mathbf{A}$ has only real eigenvalues:

- Find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of the matrix $\mathbf{A}$.
- Write down the general solution in the form

$$\mathbf{x} = C_1\mathbf{v}_1e^{\lambda_1 t} + C_2\mathbf{v}_2e^{\lambda_2 t} + \cdots + C_n\mathbf{v}_ne^{\lambda_n t},$$

where $C_1, C_2, \ldots, C_n$ are arbitrary constants.

If the matrix $\mathbf{A}$ has some complex eigenvalues (which must occur in complex conjugate pairs λ and $\overline{\lambda}$, with corresponding complex conjugate eigenvectors $\mathbf{v}$ and $\overline{\mathbf{v}}$), then a further step is necessary:

- Replace the complex terms $\mathbf{v}e^{\lambda t}$ and $\overline{\mathbf{v}}e^{\overline{\lambda}t}$ appearing in the general solution with $\mathrm{Re}(\mathbf{v}e^{\lambda t})$ and $\mathrm{Im}(\mathbf{v}e^{\lambda t})$.

  The general solution will then be real-valued for real $C_1, C_2, \ldots, C_n$.

The above procedure fails if $\mathbf{A}$ does not have n linearly independent eigenvectors. This case is not covered in this module.

4. The **complementary function** of an inhomogeneous system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h}$ is the general solution of the **corresponding homogeneous system** $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$. A **particular integral** of the inhomogeneous system is any solution of it containing no arbitrary constants.

5. The general solution of the inhomogeneous system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h}$ takes the form $\mathbf{x}_\mathrm{c} + \mathbf{x}_\mathrm{p}$, where $\mathbf{x}_\mathrm{c}$ is the complementary function and $\mathbf{x}_\mathrm{p}$ is a particular integral.

6. To find a particular integral $\mathbf{x}_\mathrm{p} = [x_\mathrm{p} \quad y_\mathrm{p}]^T$ of $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h}$, where $\mathbf{A}$ is a 2×2 matrix, try a solution constructed as follows.

   A similar procedure works for systems of more than two equations.

   - When the elements of $\mathbf{h}$ are polynomials of degree k, choose x_p and y_p to be polynomials of degree k.
   - When the elements of $\mathbf{h}$ are multiples of the same exponential function, choose x_p and y_p to be multiples of this exponential function.

   The **trial solution** will have a number of undetermined coefficients. To find their values, substitute the trial solution into the system of differential equations and equate coefficients of corresponding terms. This will give a number of equations that can be solved for the coefficients.

7. A **system of homogeneous linear constant-coefficient second-order differential equations** for unknowns $x_1, x_2, \ldots, x_n$ takes the form

$$\begin{aligned}
\ddot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\
\ddot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\
&\vdots \\
\ddot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n,
\end{aligned}$$

   where there is the same number n of equations as of unknowns.

   The equations may be expressed in matrix form as

$$\ddot{\mathbf{x}} = \mathbf{A}\mathbf{x},$$

   where $\mathbf{A}$ is a constant matrix.

8. To solve a system $\ddot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an $n \times n$ matrix with n distinct real eigenvalues, do the following.
   - Find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$, and a corresponding set of eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
   - Each *positive* eigenvalue λ, corresponding to an eigenvector $\mathbf{v}$, gives rise to two linearly independent solutions
     $$\mathbf{v}e^{\sqrt{\lambda}t} \quad \text{and} \quad \mathbf{v}e^{-\sqrt{\lambda}t}.$$
     Each *negative* eigenvalue λ, corresponding to an eigenvector $\mathbf{v}$, gives rise to two linearly independent solutions
     $$\mathbf{v}\cos\sqrt{-\lambda}t \quad \text{and} \quad \mathbf{v}\sin\sqrt{-\lambda}t.$$
     A *zero* eigenvalue corresponding to an eigenvector $\mathbf{v}$ gives rise to two linearly independent solutions
     $$\mathbf{v} \quad \text{and} \quad \mathbf{v}t.$$
   - The general solution is then an arbitrary linear combination of the $2n$ linearly independent solutions found in the previous step, involving $2n$ arbitrary real constants.

Complex eigenvalues and repeated real eigenvalues are not discussed, but they can be dealt with by generalising here.

9. Systems of second-order differential equations arise in the study of oscillating systems. A **normal mode of oscillation** is one in which all of the coordinates of the system oscillate sinusoidally with the same angular frequency.

Unit 7 Functions of several variables

1. A function $f(x, y)$ defines a surface $z = f(x, y)$, which can be visualised in a perspective view. If one of the variables is given a fixed value ($y = a$, say), then the corresponding **section function** $f(x, a)$ can be plotted as a normal graph. A **contour map** for $f(x, y)$ is a collection of **contour lines** in the xy-plane. Along each contour line, $f(x, y)$ has a constant value.

   For a function $f(x, y, z)$ of three variables, the contour lines are replaced by contour surfaces.

2. The **partial derivative** $\partial f/\partial x$ of a function $f(x, y)$ is obtained by differentiating $f(x, y)$ with respect to x, *treating y as a constant*. It is also written as $f_x(x, y)$ or f_x. All the usual rules of differentiation apply. More formally,
   $$\frac{\partial f}{\partial x} = f_x(x, y) = \lim_{\delta x \to 0} \frac{f(x + \delta x, y) - f(x, y)}{\delta x}.$$
   Similar definitions apply to $\partial f/\partial y = f_y(x, y)$, and to partial derivatives of functions of more than two variables. In a partial derivative with respect to a given independent variable, all the other independent variables are treated as constants.

In this module, unless told otherwise, you may assume that functions are 'sufficiently smooth' for their derivatives to be defined and well-behaved.

3. Functions obtained by partially differentiating a function once are called **first-order partial derivatives**. These can be partially differentiated again to obtain higher-order partial derivatives. For example,

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = f_{xx} \quad \text{and} \quad \frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = f_{yx}$$

are **second-order partial derivatives**.

4. For a sufficiently smooth function $f(x, y)$, the **mixed partial derivative theorem** states that

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x}, \quad \text{or equivalently} \quad f_{xy} = f_{yx}.$$

A similar result applies to a function $f(x_1, x_2, \ldots)$ of n variables. Again, the order of differentiation does not matter, so

$$\frac{\partial^2 f}{\partial x_i\,\partial x_j} = \frac{\partial^2 f}{\partial x_j\,\partial x_i} \quad \text{for all } x_i \text{ and } x_j.$$

5. The **chain rule of partial differentiation** takes several different forms.

- *Chain rule for small changes* If $f = f(x, y)$, and the independent variables x and y change by small amounts δx and δy, then the corresponding small change in f is

$$\delta f \simeq \frac{\partial f}{\partial x}\,\delta x + \frac{\partial f}{\partial y}\,\delta y.$$

- *Chain rule for differentiation with respect to a parameter* If $f = f(x, y)$, and the independent variables $x = x(t)$ and $y = y(t)$ are functions of a parameter t, then the rate of change of f with respect to t is

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\,\frac{dx}{dt} + \frac{\partial f}{\partial y}\,\frac{dy}{dt}.$$

- *Chain rule for a change of variables* If $f = f(x, y)$, and the independent variables $x = x(u, v)$ and $y = y(u, v)$ are functions of two other variables, u and v, then the partial derivative of f with respect to u is

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\,\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\,\frac{\partial y}{\partial u},$$

with a similar result for the partial derivative of f with respect to v.

All these chain rules can be extended to functions of more than two variables. For example, if $f = f(x, y, z)$, then the chain rule for small changes becomes

$$\delta f \simeq \frac{\partial f}{\partial x}\,\delta x + \frac{\partial f}{\partial y}\,\delta y + \frac{\partial f}{\partial z}\,\delta z.$$

6. Given a surface $z = f(x, y)$, and a direction in the xy-plane defined by a unit vector $\widehat{\mathbf{n}} = \cos\theta\,\mathbf{i} + \sin\theta\,\mathbf{j}$, the **slope of the surface** in the direction of the unit vector is

$$\text{slope} = \cos\theta\,\frac{\partial f}{\partial x} + \sin\theta\,\frac{\partial f}{\partial y} = \widehat{n}_x\, f_x + \widehat{n}_y\, f_y.$$

7. For a function $f(x, y)$, the **gradient vector** is vector-valued function of x and y in the xy-plane given by

$$\mathbf{grad}\, f = \frac{\partial f}{\partial x}\,\mathbf{i} + \frac{\partial f}{\partial y}\,\mathbf{j}.$$

It has the following properties:

- For a unit vector $\widehat{\mathbf{n}}$ in the xy-plane, the slope of the surface $z = f(x, y)$ in the direction of the unit vector is

$$\text{slope} = \widehat{\mathbf{n}} \cdot \mathbf{grad}\, f.$$

- **grad** f points in the direction in the xy-plane that gives maximum slope, and its magnitude is equal to the maximum slope.
- At any given point, **grad** f is perpendicular to the contour line through the point.

8. The **first-order Taylor polynomial** for $f(x, y)$ about (a, b) is

$$p_1(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

This matches the values and first-order partial derivatives of $f(x, y)$ at (a, b).

The plane given by

$$z = p_1(x, y)$$

is called the **tangent plane** for $f(x, y)$ at (a, b). For any direction in the xy-plane, the slopes of the function $f(x, y)$ and the corresponding tangent plane are identical at (a, b).

9. The **second-order Taylor polynomial** for $f(x, y)$ about (a, b) is

$$\begin{aligned} p_2(x, y) &= p_1(x, y) \\ &\quad + \tfrac{1}{2}\left(f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2\right). \end{aligned}$$

This matches the values and first- and second-order partial derivatives of $f(x, y)$ at (a, b).

10. A point (a, b) inside the domain of $f(x, y)$ is a **stationary point** if both $f_x(a, b)$ and $f_y(a, b)$ are equal to zero. There are three types of stationary point.

- A **local minimum** occurs at (a, b) if there is a small region around (a, b) within which $f(x, y) > f(a, b)$ at all points $(x, y) \neq (a, b)$.

- A **local maximum** occurs at (a, b) if there is a small region around (a, b) within which $f(x, y) < f(a, b)$ at all points $(x, y) \neq (a, b)$.
- A **saddle point** is a stationary point that is neither a local minimum nor a local maximum. Through such a point, some paths climb to higher function values while others descend to lower function values.

Stationary points are found by taking all the first-order partial derivatives of the given function, setting them equal to zero, and solving the resulting set of simultaneous equations. For example, the stationary points of $f(x, y)$ occur at points (x, y) that satisfy the simultaneous equations

$$f_x(x, y) = 0 \quad \text{and} \quad f_y(x, y) = 0.$$

11. The **Hessian matrix** is the matrix of the second-order partial derivatives of f.
12. To classify the stationary point at (a, b) of a smooth function $f(x, y)$, do the following.
    (a) Find the second-order partial derivatives, and evaluate them at the stationary point.
    (b) Construct the Hessian matrix $\mathbf{H}$ at the stationary point, and find its eigenvalues.
    (c) Apply the following rules:
        - If all the eigenvalues are positive, then we have a local minimum.
        - If all the eigenvalues are negative, then we have a local maximum.
        - If the eigenvalues have mixed signs, then we have a saddle point.

The test is inconclusive if any of the eigenvalues are equal to zero.

Alternatively, for a function $f(x, y)$ of two variables, the **determinant test** can be used. In this case the Hessian matrix can be written as

$$\mathbf{H} = \begin{bmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}.$$

The Hessian matrix of a smooth function is always symmetric.

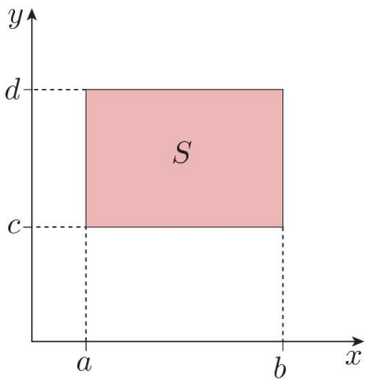
The stationary point is:

- a local minimum if $\det \mathbf{H} = AC - B^2 > 0$ and $A > 0$
- a local maximum if $\det \mathbf{H} = AC - B^2 > 0$ and $A < 0$
- a saddle point if $\det \mathbf{H} = AC - B^2 < 0$.

We could equally well use C instead of A in the determinant test.

The determinant test is inconclusive if $\det \mathbf{H} = 0$.

Unit 8 Multiple integrals



A rectangular region of integration

1. The **area integral** of a function $f(x, y)$ over a *rectangular region* S in the xy-plane, bounded by the lines $x = a$, $x = b$ and $y = c$, $y = d$, can be found by two successive integrations. The inner integral is always performed first. For example, we can write

$$\int_S f(x,y)\,dA = \int_{y=c}^{y=d} \left(\int_{x=a}^{x=b} f(x,y)\,dx \right) dy,$$

where the integral over x is carried out first, with y held constant. This gives a function of y, which is integrated over y.

Alternatively, the same area integral can be written as

$$\int_S f(x,y)\,dA = \int_{x=a}^{x=b} \left(\int_{y=c}^{y=d} f(x,y)\,dy \right) dx.$$

In this case, the integral over y is carried out first, treating x as a constant; the result is then integrated over x.

2. The **volume integral** of a function $f(x, y, z)$ over a *cuboid region* R, bounded by the planes $x = a_1$, $x = a_2$, $y = b_1$, $y = b_2$, $z = c_1$ and $z = c_2$, can be found by three successive integrations. The innermost integral is evaluated first, and successive integrations move progressively outwards. For example, we can write

$$\int_R f(x,y,z)\,dV = \int_{x=a_1}^{x=a_2} \left(\int_{y=b_1}^{y=b_2} \left(\int_{z=c_1}^{z=c_2} f(x,y,z)\,dz \right) dy \right) dx,$$

where the integral over z is done first, with x and y held constant. The result is integrated over y, with x held constant, and the final integral is over x.

Alternative orderings may be chosen. Inner integrals are always completed before integrals placed outside them. The limits of integration in a given definite integral can depend only on the variables of integration in integrals that lie further *outside* it (and are done after it). The limits of the outermost integral are always constants.

3. In an area integral, if the function to be integrated takes the **product form** $f(x, y) = g(x)\,h(y)$ *and the limits of integration are all constants*, then the area integral can be expressed as a product of definite integrals over x and y:

$$\int_S f(x,y)\,dA = \int_{x=a}^{x=b} g(x)\,dx \times \int_{y=c}^{y=d} h(y)\,dy.$$

In a volume integral, if the function to be integrated takes the product form $f(x, y, z) = u(x)\,v(y)\,w(z)$ *and the limits of integration are all constants*, then the volume integral can be expressed as a product of definite integrals over x, y and z:

$$\int_R f(x,y,z)\,dV = \int_{x=a_1}^{x=a_2} u(x)\,dx \times \int_{y=b_1}^{y=b_2} v(y)\,dy \times \int_{z=c_1}^{z=c_2} w(z)\,dz.$$

However, such factorisations do not work in general.

4. To find the area integral of a function $f(x, y)$ over any region S in the xy-plane, start by choosing which integral to do first – that over x or that over y. The following steps assume that the integration over y is done first.

   (a) Draw a diagram showing the region of integration S.

   (b) Draw a vertical strip parallel to the y-axis, centred on x, and spanning the region (as shown in the diagram). Determine the lower limit $y = \alpha(x)$ and the upper limit $y = \beta(x)$ for this strip. These are the limits for the y-integration (the inner integration). In general, they are non-constant functions of x.

   (c) Determine the minimum value $x = a$ and the maximum value $x = b$ for x-values throughout the region. These are the limits for the x-integration (the outer integration), and are always constants.

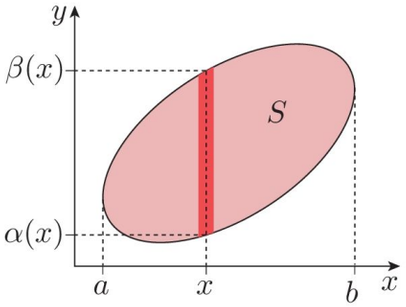
   (d) Write down the area integral as

$$\int_S f(x, y)\, dA = \int_{x=a}^{x=b} \left(\int_{y=\alpha(x)}^{y=\beta(x)} f(x, y)\, dy \right) dx.$$

   (e) Evaluate the inner integral over y first, holding x constant, and substitute in the limits of integration. This gives a function

$$g(x) = \int_{y=\alpha(x)}^{y=\beta(x)} f(x, y)\, dy.$$

   (f) Evaluate the remaining definite integral of $g(x)$ over x.



Sketching a region of integration when the integral over y is done first

   If you choose to do the x-integral first, then the limits of integration must be found using a different sketch, with horizontal strips running parallel to the x-axis. The lower and upper limits of the x-integration are then functions $x = u(y)$ and $x = v(y)$, and the lower and upper limits of the y-integration are constants $y = c$ and $y = d$. The integral is then written in the form

$$\int_S f(x, y)\, dA = \int_{y=c}^{y=d} \left(\int_{x=u(y)}^{x=v(y)} f(x, y)\, dx \right) dy.$$

   The integral over x is done first, with y held constant; the result is then integrated over y.

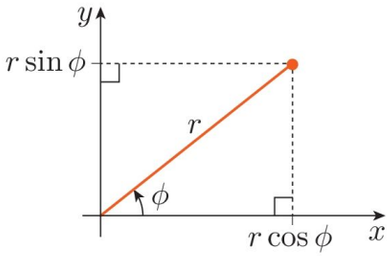
5. The volume integral of a function $f(x, y, z)$ over any region R can be written as

$$\int_R f(x, y, z)\, dV = \int_{x=a}^{x=b} \left(\int_{y=\alpha(x)}^{y=\beta(x)} \left(\int_{z=u(x,y)}^{z=v(x,y)} f(x, y, z)\, dz \right) dy \right) dx.$$

   The limits of integration can be found by drawing two diagrams – a perspective view of the region R, and a projection of R onto the xy-plane. The limits of the z-integration are found from the perspective view, by considering a thin column centred on a typical point (x, y). The limits of the y-integration (which are functions of x) are found by considering a typical thin slice parallel to the y-axis, represented by a thin strip in the projection onto the xy-plane.

The constant limits for the outer x-integration are also found from this projection. This method can be adapted for different orders of integration.

6. The area of a region in the xy-plane is equal to the area integral of the function $f = 1$ over the region. Similarly, the volume of a region in three-dimensional space is equal to the volume integral of the function $f = 1$ over the region.



Polar coordinates

7. **Polar coordinates** (r, ϕ) are related to Cartesian coordinates (x, y) by the equations

$$x = r\cos\phi, \quad y = r\sin\phi.$$

In polar coordinates an area element has area

$$\delta A = r\,\delta r\,\delta\phi.$$

The area integral of a function $f(r, \phi)$ over a disc S of radius R centred on the origin is

$$\int_S f(r,\phi)\,dA = \int_{\phi=0}^{\phi=2\pi}\left(\int_{r=0}^{r=R} f(r,\phi)\,r\,dr\right)d\phi,$$

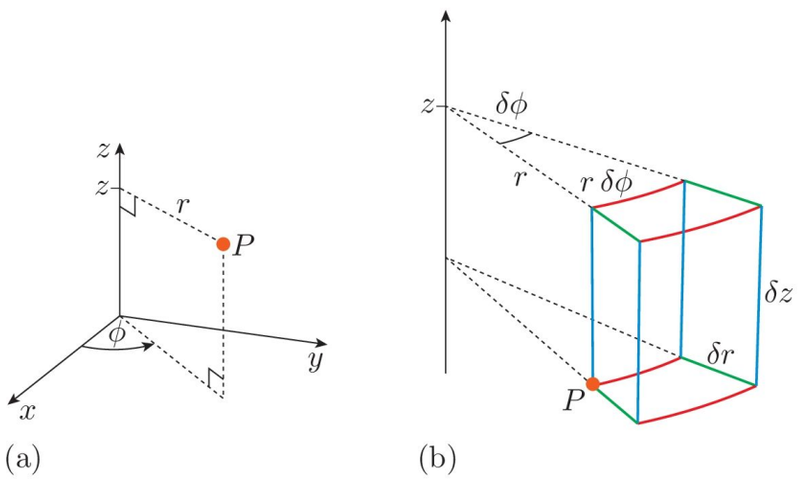
where we have chosen to integrate first over r, and then over ϕ. The reverse ordering is also valid. The limits of integration can be adjusted for regions occupying parts of a disc.

8. **Cylindrical coordinates** (r, ϕ, z) are related to Cartesian coordinates (x, y, z) by the equations

$$x = r\cos\phi, \quad y = r\sin\phi, \quad z = z.$$

In cylindrical coordinates a volume element has volume

$$\delta V = r\,\delta r\,\delta\phi\,\delta z.$$



(a) (b)

(a) Cylindrical coordinates; (b) a volume element in cylindrical coordinates

The volume integral of $f(r, \phi, z)$ over a cylindrical region D, centred on the z-axis with radius R, and with flat surfaces at $z = 0$ and $z = h$, is

$$\int_D f(r, \phi, z)\, dV = \int_{z=0}^{z=h} \left(\int_{\phi=0}^{\phi=2\pi} \left(\int_{r=0}^{r=R} f(r, \phi, z)\, r\, dr \right) d\phi \right) dz.$$

Alternative orderings are also valid. The limits of integration can be adjusted for regions based on parts of a cylinder.

9. A region that is **axially symmetric** around the z-axis has volume

$$V = \pi \int_{z=z_1}^{z=z_2} \left(r_{\max}^2(z) - r_{\min}^2(z) \right) dz,$$

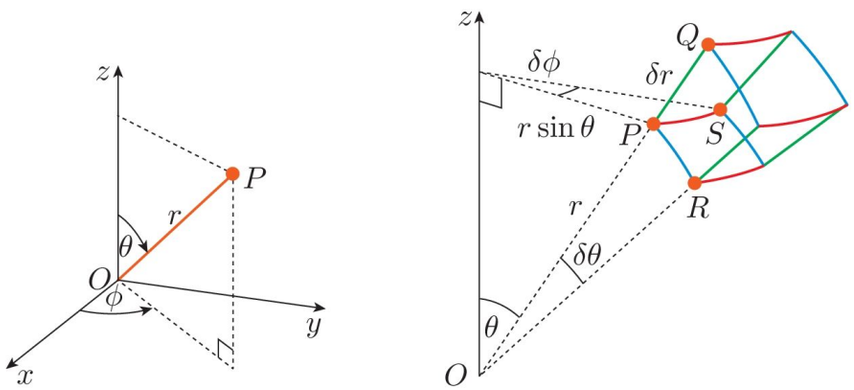
where $r_{\min}(z)$ and $r_{\max}(z)$ are the minimum and maximum radii of the region for a given value of z, and z_1 and z_2 are the minimum and maximum values of z. For a solid object with no holes, $r_{\min}(z) = 0$.

10. **Spherical coordinates** (r, θ, ϕ) are related to Cartesian coordinates (x, y, z) by the equations

$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r \cos\theta.$$

In spherical coordinates a volume element has volume

$$\delta V = r^2 \sin\theta\, \delta r\, \delta\theta\, \delta\phi.$$



(a) (b)

(a) Spherical coordinates; (b) a volume element in spherical coordinates

The volume integral of $f(r, \theta, \phi)$ over a spherical region of radius R, centred on the origin, is given by

$$I = \int_{\phi=0}^{\phi=2\pi} \left(\int_{\theta=0}^{\theta=\pi} \left(\int_{r=0}^{r=R} f(r, \theta, \phi)\, r^2 \sin\theta\, dr \right) d\theta \right) d\phi.$$

Alternative orderings are also valid. The limits of integration can be adjusted for regions based on parts of a sphere.

11. A **coordinate line** is a line along which one coordinate varies while the other coordinates remain fixed.

A coordinate system is said to be **orthogonal** if its coordinate lines, for different coordinates, meet at right angles. Cartesian, polar, cylindrical and spherical coordinate systems are all orthogonal.

12. For any coordinate u, the length of the segment of the u-coordinate line between u and $u + \delta u$, where $\delta u > 0$, is expressed as

$$\text{length of segment} = h_u\,\delta u,$$

where h_u is called the **scale factor** for the u-coordinate; this may be a function of the coordinates.

If the coordinates (u, v, w) are related to Cartesian coordinates by equations of the form

$$x = x(u, v, w), \quad y = y(u, v, w), \quad y = z(u, v, w),$$

then the scale factor h_u is given by

$$h_u = \sqrt{\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2},$$

with similar formulas for h_v and h_w.

Scale factors in some orthogonal coordinate systems are given in the table below.

Cartesian (x, y)	$h_x = 1,\ h_y = 1$
Cartesian (x, y, z)	$h_x = 1,\ h_y = 1,\ h_z = 1$
Polar (r, ϕ)	$h_r = 1,\ h_\phi = r$
Cylindrical (r, ϕ, z)	$h_r = 1,\ h_\phi = r,\ h_z = 1$
Spherical (r, θ, ϕ)	$h_r = 1,\ h_\theta = r,\ h_\phi = r\sin\theta$

In any orthogonal coordinate system (u, v) with scale factors h_u and h_v, an area element has area

$$\delta A = h_u h_v\,\delta u\,\delta v.$$

In any orthogonal coordinate system (u, v, w) with scale factors h_u, h_v and h_w, a volume element has volume

$$\delta V = h_u h_v h_w\,\delta u\,\delta v\,\delta w.$$

13. A surface can be parametrised by two coordinates (u, v). These are related to Cartesian coordinates by equations of the form

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v).$$

Using these equations, we define the **Jacobian vector**

$$\mathbf{J} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & \dfrac{\partial z}{\partial u} \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} & \dfrac{\partial z}{\partial v} \end{vmatrix},$$

which is perpendicular to the surface.

Then a surface area element has area

$$\delta A = |\mathbf{J}|\,\delta u\,\delta v,$$

where $|\mathbf{J}|$ is the magnitude of $\mathbf{J}$.

For a surface region S with $u_1 \le u \le u_2$ and $v_1 \le v \le v_2$, where the limits of u and v are all constants, the surface integral of a function $f(u, v)$ is

$$\int_S f\,dA = \int_{v=v_1}^{v=v_2} \left(\int_{u=u_1}^{u=u_2} f(u,v)\,|\mathbf{J}|\,du \right) dv.$$

For $f = 1$, this gives the area of the surface region S. If the surface S is part of the surface of the sphere of radius R centred at the origin, and (u, v) are the spherical coordinates (θ, ϕ), then $\mathbf{J} = R^2 \sin\theta\,\mathbf{e}_r$. If S is part of a cylinder of radius R with axis along the z-axis, and (u, v) are the cylindrical coordinates (ϕ, z), then $\mathbf{J} = R\mathbf{e}_r$.

Unit 9 Differentiating scalar and vector fields

1. A field is a quantity with definite values at points throughout a region of space. **Scalar fields** and **vector fields** describe the distribution of scalar and vector quantities, respectively. At any given point:
   - The value of a scalar field is independent of the orientation of the coordinate system.
   - The magnitude and direction of a vector field are independent of the orientation of the coordinate system. (This implies that the components of a vector field depend on the choice of coordinate system.)
2. In a coordinate system (u, v, w), the **unit vector** $\mathbf{e}_u$ associated with the coordinate u is a vector of unit length pointing in the direction in which u increases while the other coordinates v and w are held constant. In orthogonal coordinate systems, the unit vectors are mutually orthogonal. They may vary from point to point.

   In *Cartesian coordinates* (x, y, z), the unit vectors are

   $$\mathbf{e}_x = \mathbf{i}, \quad \mathbf{e}_y = \mathbf{j}, \quad \mathbf{e}_z = \mathbf{k}.$$

   In *polar coordinates* (r, ϕ), the unit vectors are

   $$\mathbf{e}_r = \cos\phi\,\mathbf{i} + \sin\phi\,\mathbf{j}, \quad \mathbf{e}_\phi = -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}.$$

   In *cylindrical coordinates* (r, ϕ, z), the unit vectors are

   $$\mathbf{e}_r = \cos\phi\,\mathbf{i} + \sin\phi\,\mathbf{j}, \quad \mathbf{e}_\phi = -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}, \quad \mathbf{e}_z = \mathbf{k}.$$

   In *spherical coordinates* (r, θ, ϕ), the unit vectors are

   $$\mathbf{e}_r = \sin\theta\cos\phi\,\mathbf{i} + \sin\theta\sin\phi\,\mathbf{j} + \cos\theta\,\mathbf{k},$$
   $$\mathbf{e}_\theta = \cos\theta\cos\phi\,\mathbf{i} + \cos\theta\sin\phi\,\mathbf{j} - \sin\theta\,\mathbf{k},$$
   $$\mathbf{e}_\phi = -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}.$$

3. To convert a vector field $\mathbf{F}$ from Cartesian coordinates into another orthogonal system (u, v, w), for each component do the following.

   (a) Write down

   $$F_u = \mathbf{e}_u \cdot \mathbf{F},$$

   and expand the scalar product on the right-hand side using Cartesian expressions for $\mathbf{e}_u$ and $\mathbf{F}$ (involving $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$).

   (b) The resulting expression generally depends on (x, y, z). Use the coordinate transformation equations in the form

   $$x = x(u, v, w), \quad y = y(u, v, w), \quad z = z(u, v, w),$$

   to obtain an expression for F_u solely in terms of u, v and w.

   Once this has been done for each component, write down

   $$\mathbf{F} = F_u\,\mathbf{e}_u + F_v\,\mathbf{e}_v + F_w\,\mathbf{e}_w.$$

4. Given a scalar field $V(x, y, z)$, the corresponding **gradient vector field** is

   $$\boldsymbol{\nabla} V = \frac{\partial V}{\partial x}\,\mathbf{i} + \frac{\partial V}{\partial y}\,\mathbf{j} + \frac{\partial V}{\partial z}\,\mathbf{k}.$$

   The alternative notation **grad** V is often used.

   This is also called the **gradient** of V.

5. At each point, the gradient has the following properties:

   - Its direction is that in which V increases most rapidly. This direction is perpendicular to the contour surfaces of V.
   - Its magnitude is the maximum rate of increase of V with respect to distance travelled in three-dimensional space.
   - The gradient of a scalar field is a vector field, so its magnitude and direction are independent of the orientation of the coordinate system.
   - For a small displacement $\delta\mathbf{s} = \delta x\,\mathbf{i} + \delta y\,\mathbf{j} + \delta z\,\mathbf{k}$, the corresponding small change in V is

     $$\delta V = \boldsymbol{\nabla} V \cdot \delta\mathbf{s}.$$

   - The rate of change of V with distance in the direction of the unit vector $\widehat{\mathbf{n}}$ is equal to the component of $\boldsymbol{\nabla} V$ in the direction of $\widehat{\mathbf{n}}$:

     $$\text{rate of change of } V = \widehat{\mathbf{n}} \cdot \boldsymbol{\nabla} V.$$

6. In *polar coordinates*,

   $$\boldsymbol{\nabla} V = \frac{\partial V}{\partial r}\,\mathbf{e}_r + \frac{1}{r}\frac{\partial V}{\partial \phi}\,\mathbf{e}_\phi.$$

   In *cylindrical coordinates*,

   $$\boldsymbol{\nabla} V = \frac{\partial V}{\partial r}\,\mathbf{e}_r + \frac{1}{r}\frac{\partial V}{\partial \phi}\,\mathbf{e}_\phi + \frac{\partial V}{\partial z}\,\mathbf{e}_z.$$

   In *spherical coordinates*,

   $$\boldsymbol{\nabla} V = \frac{\partial V}{\partial r}\,\mathbf{e}_r + \frac{1}{r}\frac{\partial V}{\partial \theta}\,\mathbf{e}_\theta + \frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi}\,\mathbf{e}_\phi.$$

7. Given a vector field $\mathbf{F} = F_x\,\mathbf{i} + F_y\,\mathbf{j} + F_z\,\mathbf{k}$, the corresponding **divergence** is

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

The alternative notation $\operatorname{div} \mathbf{F}$ is often used.

The divergence of a vector field $\mathbf{F}$ is a scalar field. It describes the extent to which $\mathbf{F}$ flows outwards or diverges from each point.

8. In *polar coordinates*,

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{r}\frac{\partial (rF_r)}{\partial r} + \frac{1}{r}\frac{\partial F_\phi}{\partial \phi}.$$

In *cylindrical coordinates*,

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{r}\frac{\partial (rF_r)}{\partial r} + \frac{1}{r}\frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}.$$

In *spherical coordinates*,

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{r^2}\frac{\partial (r^2 F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial (\sin\theta F_\theta)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial \phi}.$$

9. Given a vector field $\mathbf{F} = F_x\,\mathbf{i} + F_y\,\mathbf{j} + F_z\,\mathbf{k}$ expressed in a right-handed Cartesian coordinate system, the corresponding **curl** is

$$\boldsymbol{\nabla} \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}.$$

The alternative notation $\mathbf{curl}\,\mathbf{F}$ is often used.

In the determinant, the partial derivative operators in the second row act on the components in the third row. Expanding out the determinant gives

$$\boldsymbol{\nabla} \times \mathbf{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\mathbf{i} - \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z}\right)\mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\mathbf{k}.$$

The curl of a vector field $\mathbf{F}$ is another vector field. It describes the extent to which $\mathbf{F}$ rotates or swirls locally about each point. (The sense of the rotation is given by the right-hand rule.)

10. In *polar coordinates*,

$$\boldsymbol{\nabla} \times \mathbf{F} = \frac{1}{r}\left(\frac{\partial (rF_\phi)}{\partial r} - \frac{\partial F_r}{\partial \phi}\right)\mathbf{e}_z.$$

In *cylindrical coordinates*,

$$\boldsymbol{\nabla} \times \mathbf{F} = \frac{1}{r}\begin{vmatrix} \mathbf{e}_r & r\,\mathbf{e}_\phi & \mathbf{e}_z \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ F_r & rF_\phi & F_z \end{vmatrix}.$$

In *spherical coordinates*,

$$\boldsymbol{\nabla} \times \mathbf{F} = \frac{1}{r^2\sin\theta}\begin{vmatrix} \mathbf{e}_r & r\,\mathbf{e}_\theta & r\sin\theta\,\mathbf{e}_\phi \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \phi} \\ F_r & rF_\theta & r\sin\theta\,F_\phi \end{vmatrix}.$$

Unit 10 Integrating scalar and vector fields

1. A **path** is a curve with a definite sense of progression from a start point to an end point. If the start and end points are identical, then the path is closed; otherwise, it is open.

   A path can be represented by a set of **parametric equations** of the form

   $$x = x(t), \quad y = y(t), \quad z = z(t) \quad (t_1 \leq t \leq t_2),$$

   where x, y and z are Cartesian coordinates of points on the path, and the parameter t increases monotonically, from $t = t_1$ at the start point to $t = t_2$ at the end point. For a path in the xy-plane, we need only the parametric equations for x and y.

2. A path in three-dimensional space, with parametric equations

   $$x = x(t), \quad y = y(t), \quad z = z(t) \quad (t_1 \leq t \leq t_2),$$

   has **length**

   $$L = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt.$$

   If the path is restricted to the xy-plane, then this simplifies to

   $$L = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

3. The **line integral of a scalar function** $\lambda(x, y)$ along a path with parametric equations $x = x(t)$, $y = y(t)$ for $t_1 \leq t \leq t_2$ is given by

   This formula can be extended in an obvious way for paths in three-dimensional space.

   $$M = \int_{t_1}^{t_2} \lambda(x(t), y(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

   If $\lambda(x, y)$ is the mass per unit length along the path, then the line integral gives the total mass along the path.

4. To calculate the **line integral of a vector field**

   $$\mathbf{F} = F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}$$

   along a path C with the parametric representation

   $$x = x(t), \quad y = y(t), \quad z = z(t) \quad (t_1 \leq t \leq t_2),$$

   do the following.

   (a) Use the parametric representation to find

   $$\frac{d\mathbf{s}}{dt} = \frac{dx}{dt}\,\mathbf{i} + \frac{dy}{dt}\,\mathbf{j} + \frac{dz}{dt}\,\mathbf{k}.$$

   (b) Express the components of $\mathbf{F}$ as functions of the parameter t.

   (c) Find the scalar product

   $$\mathbf{F} \cdot \frac{d\mathbf{s}}{dt} = F_x \frac{dx}{dt} + F_y \frac{dy}{dt} + F_z \frac{dz}{dt}$$

   as a function of t.

(d) Evaluate the line integral as a definite integral over t:

$$\int_C \mathbf{F} \cdot d\mathbf{s} = \int_{t_1}^{t_2} \mathbf{F} \cdot \frac{d\mathbf{s}}{dt}\, dt.$$

5. A vector field $\mathbf{F}$ is called a **gradient field** if it can be expressed in the form

$$\mathbf{F} = -\boldsymbol{\nabla} U,$$

where U is a scalar field, known as the **scalar potential field** associated with $\mathbf{F}$.

Line integrals of gradient fields have the following properties:

- Any line integral of a gradient field is path-independent (i.e. it does not depend on the detailed path between given start and end points).
- Any line integral of a gradient field around a closed loop is equal to zero.

6. A vector field $\mathbf{F}$ is said to be **conservative** if, throughout its domain of definition, all its line integrals are path-independent.

The line integral of a conservative field $\mathbf{F}$ along a path C from a start point A to an end point B may be written as

$$\int_{A \to B} \mathbf{F} \cdot d\mathbf{s},$$

as there is no need to indicate the precise path C.

For a conservative field $\mathbf{F}$ with an associated scalar potential field $U(\mathbf{r})$,

$$\int_{\mathbf{r}_A \to \mathbf{r}_B} \mathbf{F} \cdot d\mathbf{s} = U(\mathbf{r}_A) - U(\mathbf{r}_B).$$

All gradient fields are conservative, and all conservative fields are gradient fields, so the terms conservative field and gradient field are synonymous, and can be used interchangeably.

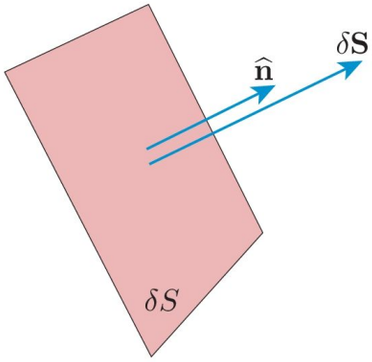
7. Given a conservative vector field $\mathbf{F}$, the corresponding scalar potential field is defined as follows. We choose a point $\mathbf{r}_0$ as the position of zero scalar potential; by definition, $U(\mathbf{r}_0) = 0$. At any point $\mathbf{r}$, the scalar potential field is then given by the line integral

$$U(\mathbf{r}) = -\int_{\mathbf{r}_0 \to \mathbf{r}} \mathbf{F} \cdot d\mathbf{s},$$

where the path may be taken to be a straight line from $\mathbf{r}_0$ to $\mathbf{r}$, or any other choice that simplifies the integration. A useful check is that $U(\mathbf{r})$ should satisfy

$$\mathbf{F} = -\boldsymbol{\nabla} U.$$

In this module all regions are simple.



An oriented area

8. The **curl test** is used to determine whether a given vector field $\mathbf{F}$ is conservative in a simple region (such as $\mathbb{R}^3$). The vector field $\mathbf{F}$ is conservative if and only if $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$ throughout the region.

9. If $\widehat{\mathbf{n}}$ is a unit vector perpendicular to a tiny planar surface element of area δS, then the **oriented area** of the element is

$$\delta\mathbf{S} = \delta S\, \widehat{\mathbf{n}}.$$

Given a vector field $\mathbf{F}$ and a planar element with oriented area $\delta\mathbf{S}$, the **flux** of the vector field over the element is defined as

$$\text{flux} = \mathbf{F} \cdot \delta\mathbf{S},$$

where the field $\mathbf{F}$ is evaluated at the position of the element. This is the normal component of the field (i.e. the component in the direction of the unit normal) multiplied by the area of the planar element. Flux is a scalar quantity that can be positive, negative or zero depending on the relative orientations of $\mathbf{F}$ and the unit normal $\widehat{\mathbf{n}}$.

10. A **closed surface** is one that divides space into two regions – an exterior space and an interior space. An **open surface** is one that is not closed. When a surface is approximated by many small planar elements, neighbouring elements are chosen to have unit normals that are almost parallel (rather than almost antiparallel). For any closed surface, all the unit normals are chosen to point outwards into the exterior space, rather than inwards towards the enclosed volume.

11. Suppose that a surface S is parametrised by (u, v), for $u_1 \leq u \leq u_2$ and $v_1 \leq v \leq v_2$, where the minimum and maximum values of the parameters are constants. Then the **flux** of a vector field $\mathbf{F}$ over a surface S is given by the surface integral

$$\text{flux} = \int_S \mathbf{F} \cdot d\mathbf{S} = \int_{v=v_1}^{v=v_2} \left(\int_{u=u_1}^{u=u_2} \mathbf{F} \cdot \mathbf{J}\, du \right) dv,$$

where

$$\mathbf{J} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & \dfrac{\partial z}{\partial u} \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} & \dfrac{\partial z}{\partial v} \end{vmatrix}$$

is the Jacobian vector in item 13 for Unit 8. To evaluate the surface integral, the integrand $\mathbf{F} \cdot \mathbf{J}$ must be expressed in terms of the parameters u and v.

On the surface of a sphere of radius R, centred on the origin and parametrised by (θ, ϕ) of spherical coordinates,

$$\mathbf{J} = R^2 \sin\theta\, \mathbf{e}_r,$$

where

$$\mathbf{e}_r = \sin\theta\cos\phi\, \mathbf{i} + \sin\theta\sin\phi\, \mathbf{j} + \cos\theta\, \mathbf{k}$$

is the radial unit vector of spherical coordinates.

12. The divergence of a vector field $\mathbf{F}$ at a given point is related to the flux of $\mathbf{F}$ over a tiny surface enclosing the point. In the limit where the surface area and its enclosed volume shrink to zero,

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{\text{flux of } \mathbf{F} \text{ over surface}}{\text{volume enclosed by surface}}.$$

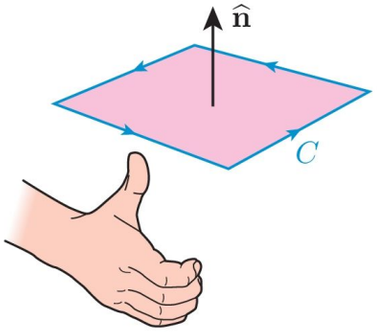
So the divergence of a vector field at any point can be interpreted as the *flux per unit volume* at that point.

13. Given a vector field $\mathbf{F}$ and a closed surface S enclosing a volume V, the **divergence theorem** states that

$$\int_S \mathbf{F} \cdot d\mathbf{S} = \int_V \boldsymbol{\nabla} \cdot \mathbf{F}\, dV.$$

This is also called **Gauss's theorem**.

14. The unit normal of a planar element and the sense of positive progression around the perimeter of the element are related by the **right-hand grip rule**: with the thumb of your right hand pointing in the direction of the unit normal of a planar element, the curled fingers of your right hand indicate the sense of positive progression around the perimeter of the element.



The right-hand grip rule

15. Given a vector field $\mathbf{F}$ and a closed path C, the **circulation** of $\mathbf{F}$ around C is given by the line integral

$$\text{circulation} = \oint_C \mathbf{F} \cdot d\mathbf{s}.$$

The circle on the integral sign is optional: it is used to indicate that the path is closed. If C is a closed path around a planar element with a given unit normal, then it is understood that C is traversed in the positive sense determined by the right-hand grip rule.

16. Given a vector field $\mathbf{F}$ in the vicinity of a given point, the component of $\boldsymbol{\nabla} \times \mathbf{F}$ in the direction of the unit vector $\hat{\mathbf{n}}$ can be found by taking a planar element with unit normal $\hat{\mathbf{n}}$ at the point. In the limit where the element becomes very small,

$$(\boldsymbol{\nabla} \times \mathbf{F}) \cdot \hat{\mathbf{n}} = \frac{\text{circulation around perimeter of element}}{\text{area of element}}.$$

So each component of the curl at a given point can be interpreted as a *circulation per unit area* at that point.

17. If $\mathbf{F}$ is a vector field and S is an open surface with perimeter C, then the **curl theorem** states that

$$\oint_C \mathbf{F} \cdot d\mathbf{s} = \int_S (\boldsymbol{\nabla} \times \mathbf{F}) \cdot d\mathbf{S},$$

where it is assumed that C is traversed in the sense determined by the right-hand grip rule. This result is also called **Stokes's theorem**.

Unit 11 Fourier series

1. A function $f(t)$ is **periodic** if there is some positive number λ such that $f(t+\lambda) = f(t)$ for all t. The number λ is a **period** of the function f. If λ is a period of a function, then so is $n\lambda$ for any positive integer n. The **fundamental period** of a periodic function is the smallest (positive) value for the period. The symbol τ is often used for the fundamental period. Unless it is specified otherwise, assume that 'period' means 'fundamental period'. A **fundamental interval** for a periodic function is any interval whose length is the fundamental period.
2. The function $f(t) = \cos(\omega t)$ is periodic with fundamental period $2\pi/\omega$. The constant ω is called the **angular frequency**.
3. A function $f(t)$ is **even** if

   $$f(-t) = f(t) \quad \text{for all values of } t.$$

   For any integer k, $\cos(kt)$ is even.

   A function $f(t)$ is **odd** if

   $$f(-t) = -f(t) \quad \text{for all values of } t.$$

   For any (non-zero) integer k, $\sin(kt)$ is odd.
4. Even and odd functions combine under addition and multiplication as follows.
   - The sum of two even functions is even.
   - The sum of two odd functions is odd.
   - The product of two even functions is even.
   - The product of two odd functions is even.
   - The product of an even and an odd function is odd.
5. The integrals of even and odd functions can be simplified when the range of integration is symmetric.
   - If $f(t)$ is even, then

     $$\int_{-a}^{a} f(t)\,dt = 2\int_{0}^{a} f(t)\,dt.$$

   - If $f(t)$ is odd, then

     $$\int_{-a}^{a} f(t)\,dt = 0.$$

6. A **Fourier series** is an infinite series of trigonometric functions of the form

$$A_0 + A_1 \cos\left(\frac{2\pi t}{\tau}\right) + A_2 \cos\left(\frac{4\pi t}{\tau}\right) + \cdots$$
$$+ B_1 \sin\left(\frac{2\pi t}{\tau}\right) + B_2 \sin\left(\frac{4\pi t}{\tau}\right) + \cdots,$$

where the constants $A_0, A_1, A_2, \ldots, B_1, B_2, \ldots$ are called the **Fourier coefficients**, and $\tau > 0$.

A Fourier series defines a periodic function whose fundamental period is τ.

7. To find the **Fourier series for a periodic function** $f(t)$, do the following.

   (a) Find the fundamental period τ.

   (b) Write down the Fourier series

$$F(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos\left(\frac{2n\pi t}{\tau}\right) + \sum_{n=1}^{\infty} B_n \sin\left(\frac{2n\pi t}{\tau}\right),$$

   where A_0 and the A_n and B_n are coefficients to be determined. Simplify the arguments of the sines and cosines where possible.

   (c) Use the following formulas to determine the Fourier coefficients:

$$A_0 = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} f(t)\,dt,$$
$$A_n = \frac{2}{\tau}\int_{-\tau/2}^{\tau/2} f(t)\cos\left(\frac{2n\pi t}{\tau}\right)dt \quad (n = 1, 2, \ldots),$$
$$B_n = \frac{2}{\tau}\int_{-\tau/2}^{\tau/2} f(t)\sin\left(\frac{2n\pi t}{\tau}\right)dt \quad (n = 1, 2, \ldots).$$

   (d) If desired, express the final Fourier series in a compact form with general formulas for its coefficients.

8. To find the **Fourier series for an odd periodic function** $f(t)$, do the following.

   (a) Identify $f(t)$ as being odd, and find its fundamental period τ.

   (b) Write down the Fourier series

$$F(t) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{2n\pi t}{\tau}\right).$$

   (c) Find the coefficients by evaluating the definite integrals

$$B_n = \frac{4}{\tau}\int_{0}^{\tau/2} f(t)\sin\left(\frac{2n\pi t}{\tau}\right)dt.$$

   (d) If desired, express the final Fourier series in a compact form with general formulas for its coefficients.

9. To find the **Fourier series for an even periodic function** $f(t)$, do the following.
   (a) Identify $f(t)$ as being even, and find its fundamental period τ.
   (b) Write down the Fourier series
   $$F(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos\left(\frac{2n\pi t}{\tau}\right).$$
   (c) Find the coefficients by evaluating the definite integrals
   $$A_0 = \frac{2}{\tau}\int_0^{\tau/2} f(t)\,dt,$$
   $$A_n = \frac{4}{\tau}\int_0^{\tau/2} f(t)\cos\left(\frac{2n\pi t}{\tau}\right)dt.$$
   (d) If desired, express the final Fourier series in a compact form with general formulas for its coefficients.
10. If $f(t)$ is a periodic function, then any interval of length equal to its fundamental period τ can be used to evaluate the Fourier coefficients.
11. The following identities are often useful when evaluating Fourier series: if n is an integer, then
    $$\cos(n\pi) = (-1)^n,$$
    $$\sin(n\pi) = 0,$$
    $$\cos\left(\frac{n\pi}{2}\right) = \begin{cases} (-1)^{n/2} & \text{for } n \text{ even}, \\ 0 & \text{for } n \text{ odd}, \end{cases}$$
    $$\sin\left(\frac{n\pi}{2}\right) = \begin{cases} 0 & \text{for } n \text{ even}, \\ (-1)^{(n+3)/2} & \text{for } n \text{ odd}. \end{cases}$$
12. The following integrals may be useful for evaluating Fourier coefficients:
    $$\int t\sin(at)\,dt = \frac{1}{a^2}\left[\sin(at) - at\cos(at)\right],$$
    $$\int t\cos(at)\,dt = \frac{1}{a^2}\left[\cos(at) + at\sin(at)\right].$$
13. The **even extension** $f_{\text{even}}(t)$ of a function $f(t)$ defined on an interval $[0, T]$ is the even periodic function with period $2T$ and fundamental interval $[-T, T]$ defined by
    $$f_{\text{even}}(t) = \begin{cases} f(t) & \text{for } 0 \le t \le T, \\ f(-t) & \text{for } -T < t < 0, \end{cases}$$
    $$f_{\text{even}}(t + 2T) = f_{\text{even}}(t).$$

14. The **odd extension** $f_{\text{odd}}(t)$ of a function $f(t)$ defined on an interval $[0, T]$ is the odd periodic function with period $2T$ and fundamental interval $[-T, T]$ defined by

$$f_{\text{odd}}(t) = \begin{cases} f(t) & \text{for } 0 \leq t \leq T, \\ -f(-t) & \text{for } -T < t < 0, \end{cases}$$

$$f_{\text{odd}}(t + 2T) = f_{\text{odd}}(t).$$

15. If $f(t)$ is a periodic function with Fourier series $F(t)$, then F **converges** to f everywhere if f is continuous. If $f(t)$ has a discontinuity at $t = t_0$, then

$$F(t_0) = \tfrac{1}{2}[f(t_0^+) + f(t_0^-)],$$

where $f(t_0^+)$ is the limit of $f(t)$ as t approaches t_0 from above, and $f(t_0^-)$ is the limit of $f(t)$ as t approaches t_0 from below.

16. If a continuous periodic function $f(t)$ with fundamental period τ has Fourier series $F(t)$, then its derivative $f'(t)$ has the same fundamental period τ, and its Fourier series is given by $F'(t)$.

Unit 12 Partial differential equations

1. A **partial differential equation** is an equation relating a dependent variable and two or more independent variables through the partial derivatives of the dependent variable. The **order** of a partial differential equation is the order of the highest derivative that occurs in it. A partial differential equation is **homogeneous** if there are no terms that are solely functions of the independent variables. It is **linear** if the terms that contain the dependent variable are proportional to the dependent variable or to one of its partial derivatives.

   Examples of partial differential equations are the wave equation and the diffusion equation.

   The one-dimensional form of the **wave equation** is

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2},$$

   where c is a positive constant called the **wave speed**.

   The one-dimensional **diffusion equation** is

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2},$$

   where D is a positive constant called the **diffusion coefficient**.

   The diffusion equation is also known as the **heat equation**.

2. Homogeneous linear partial differential equations satisfy the **principle of superposition**: if $u_1(x,t)$ and $u_2(x,t)$ both satisfy a homogeneous linear partial differential equation, then any linear combination

$$u(x,t) = a_1\,u_1(x,t) + a_2\,u_2(x,t)$$

(where a_1 and a_2 are constants) also satisfies the partial differential equation.

3. In order to find a unique solution of a partial differential equation for $u(x,t)$, we must specify boundary conditions and initial conditions.

A **boundary condition** is an equation relating to a specific value of x that holds for all time. For example, if an end of a taut string is held at a fixed position or a rod is held at a fixed temperature, then the boundary condition is $u(L,t) = 0$ for all t. If the end of a rod is thermally isolated, then the boundary condition is

$$\frac{\partial u}{\partial x}(0,t) = 0 \quad \text{for all } t.$$

An **initial condition** is an equation that relates to a specific value of t (usually $t = 0$) and specifies $u(x,t)$ at this time. For example, if the initial shape of a taut string (or the initial temperature profile of a rod) is given by $f(x)$ for $0 \leq x \leq L$, then the initial condition is $u(x,0) = f(x)$ for $0 \leq x \leq L$. If the initial velocity profile of a taut string is $g(x)$ for $0 \leq x \leq L$, then the initial condition is

$$\frac{\partial u}{\partial t}(x,0) = g(x) \quad \text{for } 0 \leq x \leq L.$$

4. To use the **method of separation of variables** to solve a linear homogeneous partial differential equation with dependent variable u and independent variables x and t, subject to given boundary and initial conditions, do the following.

   (a) Write the unknown function $u(x,t)$ as a product of functions of one variable:

   $$u(x,t) = X(x)\,T(t).$$

   Find the required partial derivatives of u in terms of the ordinary derivatives of the functions X and T. For example, $\partial^2 u/\partial x^2 = X''(x)\,T(t)$.

   (b) Substitute the partial derivatives found in step (a) into the partial differential equation. Rearrange the equation so that each side consists of a function of a single independent variable. Equate each side of the rearranged equation to the same **separation constant** μ, and hence obtain ordinary differential equations for X and T.

   For example, the wave equation separates to give the equations

   $$\frac{X''(x)}{X(x)} = \mu, \quad \frac{1}{c^2}\frac{\ddot{T}(t)}{T(t)} = \mu.$$

(c) Use the given boundary conditions for u to find boundary conditions for X.

For example, the boundary condition $u(0,t) = 0$ for all t will give the boundary condition $X(0) = 0$, and the boundary condition $(\partial u/\partial x)(L,t) = 0$ will give the boundary condition $X'(L) = 0$.

(d) Solve the differential equation for X, and apply the boundary conditions. Consider different choices for the separation constant μ. (Typically, the solutions $X(x)$ take a different form depending on whether the separation constant is positive, negative or zero.) The boundary conditions generally produce a discrete set of solutions $X_n(x)$, which are called **eigenfunctions**, and a corresponding discrete set of values μ_n for the separation constant, which are called **eigenvalues**.

(e) For each allowed μ_n, determine the corresponding solution $T_n(t)$ of the differential equation for $T(t)$.

(f) Combine $X_n(x)$ and $T_n(t)$ to obtain a family of product solutions

$$u_n(x,t) = X_n(x)\,T_n(t), \quad n = 1, 2, 3, \ldots.$$

Express the general solution as an infinite linear combination of these product solutions:

$$u(x,t) = \sum_{n=1}^{\infty} a_n\, u_n(x,t).$$

(g) Use the initial conditions and results about Fourier series to determine (when possible) the coefficients a_n.

For example, the general solution for a taut string released from rest is

$$\sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{n\pi ct}{L}\right),$$

and if the initial profile of the string is given by the initial condition $u(x,0) = f(x)$ for $0 \le x \le L$, then the coefficients A_n are given by the coefficients of the odd periodic extension of $f(x)$, namely

$$A_n = \frac{2}{L}\int_0^L \sin\left(\frac{n\pi x}{L}\right) f(x)\,dx, \quad n = 1, 2, 3, \ldots.$$

5. The general solution of the equation $X''(x) = \mu\, X(x)$ splits into three cases depending on whether μ is positive, negative or zero (all cases have two constants A and B).

- If μ is positive, then $\mu = c^2$ for some $c > 0$, and the general solution is $X(x) = A\exp(cx) + B\exp(-cx)$.
- If μ is negative, then $\mu = -k^2$ for some $k > 0$, and the general solution is $X(x) = A\sin(kx) + B\cos(kx)$.
- If μ is zero, then the general solution is $X(x) = Ax + B$.

Unit 13 Non-linear differential equations

1. A **non-linear differential equation**, or system of differential equations, is one that is not linear in at least one of the dependent variables. An **autonomous differential equation**, or system of differential equations, is one where the independent variable does not appear explicitly.

   This unit is mainly concerned with systems of two coupled autonomous differential equations, of the form

   $$\frac{dx}{dt} = u(x, y), \quad \frac{dy}{dt} = v(x, y).$$

   The system is non-linear if either u or v is a non-linear function.

2. For systems of two coupled differential equations, the solution at time t can be represented by a point $\mathbf{x}$ in a two-dimensional space called **phase space**, with coordinates x, y.

   The **phase point** $\mathbf{x}$ has equation of motion

   $$\dot{\mathbf{x}} = \mathbf{u}(x, y),$$

   where $\mathbf{u}(x, y)$ is a vector field

   $$\mathbf{u} = [u(x, y) \quad v(x, y)]^T.$$

   A particular solution $[x(t) \quad y(t)]^T$ of the equation of motion determines a path in the xy-plane, parametrised by t, whose tangent vector at any point (x, y) on it is the vector $\mathbf{u}(x, y)$. Such a solution curve is called a **phase path** (or **phase trajectory**). The xy-plane containing the solution curves is called the **phase plane**, and a diagram showing the phase paths is called a **phase diagram**.

3. An **equilibrium point** (or **fixed point**) of a system of differential equations $[\dot{x} \quad \dot{y}]^T = \mathbf{u}(x, y)$ is a point (x_e, y_e) such that $x(t) = x_e$, $y(t) = y_e$ is a constant solution of the system of differential equations; that is, $\dot{x}(t) = 0$ and $\dot{y}(t) = 0$ at the point (x_e, y_e).

   To find the equilibrium points of $\dot{\mathbf{x}} = \mathbf{u}(x, y)$ for a given vector field $\mathbf{u}$, solve the equation

   $$\mathbf{u}(x, y) = \mathbf{0}$$

   for x and y.

4. An equilibrium point of a system of differential equations is said to be **stable** when *all* points in the neighbourhood of the equilibrium point remain in the neighbourhood of the equilibrium point as time increases, and **unstable** otherwise.

5. The matrix

   $$\mathbf{J}(x, y) = \begin{bmatrix} \dfrac{\partial u}{\partial x}(x, y) & \dfrac{\partial u}{\partial y}(x, y) \\ \dfrac{\partial v}{\partial x}(x, y) & \dfrac{\partial v}{\partial y}(x, y) \end{bmatrix}$$

   is the **Jacobian matrix** of the vector field $\mathbf{u} = [u \quad v]^T$.

6. Suppose that the system of differential equations $\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x})$ has an equilibrium point at $x = x_e$, $y = y_e$. To linearise the system near this equilibrium point, do the following.

   (a) Find the Jacobian matrix $\mathbf{J}(x, y)$.

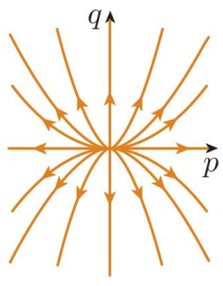
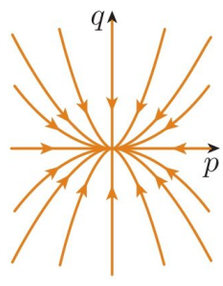
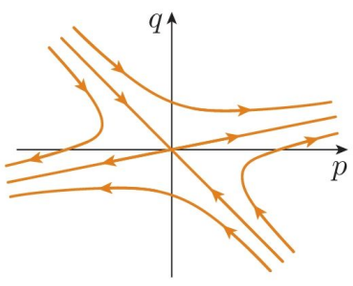
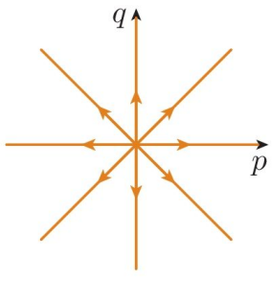
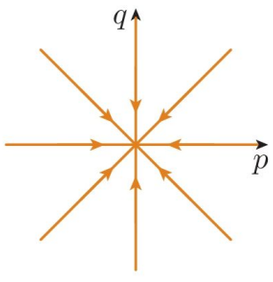
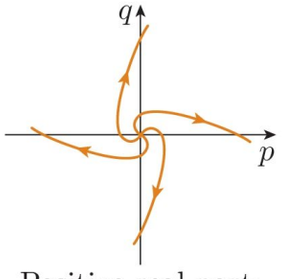
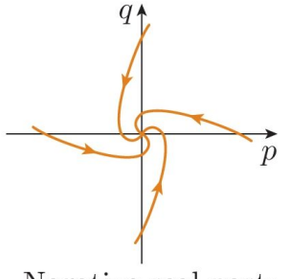
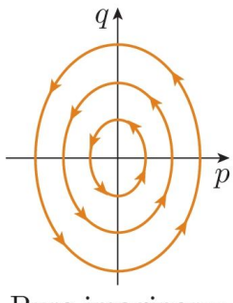
   (b) In the neighbourhood of the equilibrium point, the differential equations can be approximated by the **linearised form**

$$\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial u}{\partial x}(x_e, y_e) & \dfrac{\partial u}{\partial y}(x_e, y_e) \\ \dfrac{\partial v}{\partial x}(x_e, y_e) & \dfrac{\partial v}{\partial y}(x_e, y_e) \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix},$$

   where $x(t) = x_e + p(t)$ and $y(t) = y_e + q(t)$. The vector $[p \quad q]^T$ is the **perturbation** from the equilibrium point.

7. Consider a linear system of differential equations $\dot{\mathbf{p}} = \mathbf{A}\mathbf{p}$. This system has a single equilibrium point, at the origin. The nature of the equilibrium point is determined by the eigenvalues and eigenvectors of the matrix $\mathbf{A}$, as indicated by the following table.

There are special cases not considered in this module (such as when an eigenvalue is zero).

<table>
<tr><td>Real eigenvalues, distinct</td><td> 
Both positive: source</td><td> 
Both negative: sink</td><td> 
Differing signs: saddle</td></tr>
<tr><td>Real eigenvalues, equal</td><td colspan="2"> 
Positive: star source</td><td> 
Negative: star sink</td></tr>
<tr><td>Complex eigenvalues</td><td> 
Positive real part: spiral source</td><td> 
Negative real part: spiral sink</td><td> 
Pure imaginary: centre</td></tr>
</table>

8. The equilibrium point is **stable** if it is a sink, a star sink, a spiral sink or a centre, and **unstable** if it is a source, a star source, a spiral source or a saddle point.

9. To classify the equilibrium points of the non-linear system of differential equations $\dot{\mathbf{x}} = \mathbf{u}(x, y)$, do the following.

   (a) Find the equilibrium points using the method described in item 3.

   (b) Use the procedure described in item 6 to find the linear system that approximates the non-linear system in the neighbourhood of each equilibrium point.

   (c) Use the table in item 7 to classify the linear system, with the matrix $\mathbf{A}$ in item 7 equal to the linear approximation $\mathbf{J}(x_\mathrm{e}, y_\mathrm{e})$ found in item 6.

   The behaviour of a system of non-linear differential equations near an equilibrium point is the same as the behaviour of the linear approximation in the neighbourhood, *except* when the linear system has a centre. If the linear system has a centre, then the equilibrium point of the original non-linear system may be a stable centre, a stable spiral sink or an unstable spiral source.

10. Some equations of motion have a **constant of motion** $K(x, y)$. This function has the property that $K(x(t), y(t))$ remains constant along any given phase path $(x(t), y(t))$. This implies that K satisfies a differential equation

$$\frac{\partial K}{\partial x}\frac{dx}{dt} + \frac{\partial K}{\partial y}\frac{dy}{dt} = 0.$$

    When there is a constant of motion, the phase paths must be contours of $K(x, y)$. If the initial condition lies on a closed contour of $K(x, y)$, then the motion will be periodic in t.

11. An important application is **population dynamics**, where $x(t)$ and $y(t)$ are populations of two interacting species (usually predator and prey). The **Lotka–Volterra equations** are an important model:

$$\frac{dx}{dt} = kx\left(1 - \frac{y}{Y}\right), \quad \frac{dy}{dt} = -hy\left(1 - \frac{x}{X}\right),$$

    where k, h, X and Y are positive constants.

    The equilibrium points of the Lotka–Volterra equations are an unstable saddle at $(0, 0)$ and a stable centre at (X, Y).

    The Lotka–Volterra equations have a constant of motion

$$K(x, y) = h\ln x + k\ln y - \frac{h}{X}x - \frac{k}{Y}y.$$

    The contours are all closed curves (except for the x- and y-axes).

12. A second-order differential equation $\ddot{x} = f(x, \dot{x})$ can be converted into a pair of simultaneous first-order equations by setting $y = \dot{x}$; the equivalent pair is

$$\dot{x} = y, \quad \dot{y} = f(x, y).$$

13. The **undamped pendulum** satisfies the differential equations

$$\dot{x} = y, \quad \dot{y} = -\omega^2 \sin x \quad (-\pi < x \leq \pi).$$

This system has two equilibrium points: a stable centre at $(0, 0)$ and an unstable saddle at $(\pi, 0)$.

14. The **damped pendulum** satisfies the differential equations

$$\dot{x} = y, \quad \dot{y} = -\omega^2 \sin x - \varepsilon y \quad (-\pi < x \leq \pi),$$

where ε is a positive constant. This system has two equilibrium points, at $(0, 0)$ and $(\pi, 0)$. The equilibrium point $(0, 0)$ is a stable spiral sink for $0 < \varepsilon < 2\omega$, a stable improper sink for $\varepsilon = 2\omega$, and a stable sink for $\varepsilon > 2\omega$. The equilibrium point $(\pi, 0)$ is an unstable saddle.

Index